Science
AAAS

Supplementary Materials for

Arctic-Adapted Dogs Emerged at the Pleistocene-Holocene Transition

Authors: Mikkel-Holger S. Sinding*§, Shyam Gopalakrishnan*, Jazmín Ramos-Madrigal*, Marc de Manuel*, Vladimir V. Pitulko*, Lukas Kuderna, Tatiana R. Feuerborn, Laurent A. F. Frantz, Filipe G. Vieira, Jonas Niemann, Jose A. Samaniego Castruita, Christian Carøe, Emilie U. Andersen-Ranberg, Peter D. Jordan, Elena Y. Pavlova, Pavel A. Nikolskiy, Aleksei K. Kasparov, Varvara V. Ivanova, Eske Willerslev, Pontus Skoglund, Merete Fredholm, Sanne Eline Wennerberg, Mads Peter Heide-Jørgensen, Rune Dietz, Christian Sonne, Morten Meldgaard, Love Dalén, Greger Larson, Bent Petersen, Thomas Sicheritz-Pontén, Lutz Bachmann, Øystein Wiig, Tomas Marques-Bonet‡§, Anders J. Hansen‡§, M. Thomas P. Gilbert‡§.

*These authors contributed equally to this work.
‡These authors co-supervised this work.
§Corresponding author. Email: mhssinding@gmail.com (M.-H.S.S.); tomas.marques@upf.edu (T.M.-B.); ajhansen@sund.ku.dk (A.J.H.); tgilbert@sund.ku.dk (M.T.P.G.).

This PDF file includes:

Table of contents

Materials and Methods

Modern Greenland dog samples

A total of ten modern biological samples of tissue or DNA extract was obtained from biobanks at the Greenland Institute of Natural Resources, Institute of Bioscience at Aarhus University or Department of Veterinary and Animal Sciences at the University of Copenhagen. Specifically, these ten samples consist of two samples from each of five Greenland cities or villages, specifically, Qaanaaq, Ilulissat, Aasiaat, Sisimiut from West Greenland and Tasiilaq, East Greenland.

The Zhokhov site, background and sample

The site has probably been a base-camp at the centre of a vast mobility network in which dog sledding would have been useful. Several wooden remains resembling dog sled parts have been excavated from the same age context at the Zhokhov site (Fig. S1). The Zhokhov dog specimen (Zh-03-97) ((CGG6)) was excavated at the Zhokhov site, Zhokhov Island, New Siberian Islands in 2003 by Vladimir Pitulko. This specimen of a left mandible (Fig. S2) was directly dated to 8529 ± 30 ^{14}C years before present (MAMS-24239, calibrated to ca. 9514.5 years before present, using OxCal v4.2.4. (*27*)).

The Yana site, background and sample

The Yana (Y06-NP-18994) ((CGG23)) left mandible specimen (Fig. S3) was excavated from the Yana site by Vladimir Pitulko and Elena Y. Pavlova in 2008-2012, and collected directly from permafrost deposits. This specimen was directly dated to 28,830 ± 130 ^{14}C years before present (MAMS-24246, calibrated to ca. 33,019.5 years before present, using OxCal v4.2.4. (*27*)).

Methods and analysis

DNA extraction and sequencing

All DNA pre-amplification work on the Zhokhov and Yana samples was performed in ancient DNA facilities at the Centre for GeoGenetics, University of Copenhagen (Denmark) following ancient DNA guidelines (*28*). For ancient DNA extraction, a piece of a tooth from each specimen was crushed and then extracted in a urea-proteinase K buffer following (*29*). DNA in the digest was bound to a MinElute column (Qiagen, Hilden, Germany) in combination with a binding apparatus as described in (*30*) using a binding buffer following (*31*) in a sample:buffer ratio of 1:10. The column was subsequently treated according to the manufacturer's guidelines. Modern DNA extraction was performed using the Qiagen Blood and tissue kit (Qiagen) according to the manufacturer's guidelines. Modern DNA samples were fragmented prior to library preparation using a Bioruptor NGS device (Diagenode, Liège, Belgium). All extracted DNA was converted into double-stranded sequencing libraries using the NEBNext DNA Sample Prep Master Mix Set 2 (E6070 - New England Biolabs Inc., Beverly, MA, USA) and the "single-tube" library building protocol BEST as described in (*32*). Libraries were sequenced using Illumina HiSeq 2000 and 2500 (Illumina, San Diego, CA, USA) platforms at the National High-throughput Sequencing Center, Copenhagen, Denmark and BGI-Europe, Copenhagen, Denmark.

Quality control and alignment

We used the PALEOMIX (*33*) pipeline to process short reads obtained for all ancient and modern samples, including the samples that were obtained from previously published studies. As the first step in the pipeline, we trimmed the reads and removed adapters using

AdapterRemoval2 (*34*). Paired-end reads overlapping more than 10 base pairs - calculated using the sequences at the 3' end of the first read and the 5' end of the second read of the pair - were merged into a single long read (--collapse option). Reads shorter than 25 bp (after removing adapter sequences) were discarded. These processed reads were mapped against the wolf reference genome (*35*) and to the dog reference genome (CanFam3.1) using the alignment tool, bwa aln (v0.7.15; aln algorithm) (*36*), or as in Leathlobhair *et al*. (*3*) (specifically for the CTVT work). Duplicate reads and reads that mapped to multiple locations in the reference genome were discarded using the Picard set of tools (v1.128, https://broadinstitute.github.io/picard). In order to improve the local mapping of reads that span indels, we used GATK (v3.8.0) (*37*, *38*) to perform an indel realignment step on the mapped reads for each of the samples. Since we do not have a set of curated indels in the species of interest for these samples, we performed the indel realignment step using no external indel database. All analyses were performed using the alignments against the wolf reference genome (*35*), unless stated otherwise. The wolf reference genome was used in order to avoid potential reference biases when comparing the ancient Zhokhov sample to present-day dog genomes (*5*).

Assessing DNA damage patterns

We used mapDamage (v2.0.6) (*39*) to assess the type specific error rates estimated and aDNA damage patterns in the two ancient samples sequenced in this study. Mapped reads from the Zhokhov and Yana samples showed an increased proportion of C to T and G to A substitutions at the 5' and 3' read ends, respectively (Fig. S4). Additionally, mapDamage (v2.0.6) (*39*) was used to rescale the quality scores of bases inferred to be affected by DNA damage in the Zhokhov and Yana samples, as well as ancient reference samples, viz., the ~34,900 years BP Siberian Taimyr wolf (*5*), the ~4800 years BP Irish Newgrange dog (*40*), the ~4700 years BP German Cherry Tree Cave dog (*41*), the ~7000 years BP German Herxheim dog (*41*), the ~4200 years BP Newfoundland Port au Choix dog (*3*), the ~1100 years BP Alaskan Uyak dog (*3*), the ~960 years BP Virginian Weyanoke Old Town 1 dog (*3*) and the ~1100 years BP Virginian Weyanoke Old Town 2 dog (*3*).

Error rates

To further assess the quality of the aDNA data, we estimated type specific error rates using ANGSD (v0.921) as described in Orlando, *et al.* (*28*) for the Zhokhov and Yana samples. Majority count consensus sequences for the Andean fox and Greenland wolf genomes were used as the ancestral and perfect genome, respectively. In each case, the consensus sequence was created using ANGSD at sites with a minimum depth of coverage of 3, reads with minimum mapping quality of 30 and bases with minimum base quality of 20. Figure S5 shows the error rates estimated for Zhokhov and Yana samples as well as for reference ancient samples (Newgrange, Cherry Tree Cave dog, Herxheim dog, 3 American pre-contact dogs and Taimyr wolf) and five modern samples used in this study, shown as comparison. Overall, Zhokhov and Yana genomes display error rates comparable to those of other sequenced ancient dog and wolf genomes (Fig. S5). These errors are mostly due to increased C to T and G to A substitutions caused by aDNA damage. To account for these errors specific to ancient samples, we restricted the analysis described below to transversions, except when using called genotypes.

Genotype calling
We performed genotype calling on each of the ancient and modern samples included in the study independently, using the HaplotypeCaller algorithm implemented in GATK v3.8 (*37*). We called genotypes for each sample filtering out bases with base quality less than 20 and reads with mapping quality less than 30. This minimally filtered set of genotypes was used for Population Branch Statistics (PBS) and pairwise F_{st} analyses, with further filtering being performed in each analysis.

Reference datasets used
We used a combination of previously published whole genome sequences and genome wide SNP chip data as reference material. References and details for whole genome data are given in Table S1 and for genome wide SNP chip data is given in Table S3.

Principal Component Analysis
We performed a principal component analysis (PCA) using smartpca (*42*, *43*), from the eigensoft suite of tools (v7.2.1, https://github.com/DReichLab/EIG), to explore the affinity between Zhokhov and Yana genomes to a dataset of ancient and present-day dogs, present-day wolves and the Taimyr wolf (Table S1). For each sample and each site sampled one allele by creating a majority count consensus sequence. This consensus was created using ANGSD (v0.921) in sites with a minimum depth of coverage of 3, reads with a minimum mapping quality of 30 and sites with a minimum quality of 20. This approach allowed us to incorporate samples with heterogeneous depths of coverage. Sites with a missingness greater than 20% and minor allele frequency (MAF) < 0.05 were excluded. Additionally, to reduce the bias introduced by aDNA damage, we only included transversion in the analysis. The final dataset consisted of 2,200,623 sites. Smartpca was run using the *lsqproject* option to be able to include low coverage, high missingness samples in this analysis.

Pairwise distances
Pairwise distances between the Zhokhov genome and the dog samples in the whole-genome panel (Table S1) were calculated as the fraction differences between pairs of samples using PLINK 1.9 (*44*) and a random allele for each individual. A similar sampling and filtering approach and dataset as the one described for the PCA were used.

Admixture analyses using whole genome data (and Genotype likelihoods)
Identifying variant sites and calling genotype at these sites has been shown to introduce biases when the sequencing data used for such analyses is very heterogeneous (*45*). In order to avoid these issues, we performed a subset of the analyses using randomly sampled bases in some cases, and genotype likelihoods in others. The genotype likelihoods at variant sites were computed in ANGSD (v0.921) (*46*) using the aligned reads obtained from PALEOMIX, under the model proposed in *samtools* (v1.2) (*36*). Nucleotides with base qualities lower than 20 and reads with a mapping quality lower than 20 were discarded. Sites with data at fewer than 95 out of the 99 samples were excluded. Finally, only sites with an estimated MAF greater than 0.05 were retained. We used NGSAdmix (*47*) on the genotype likelihoods computed in ANGSD to estimate the ancestry clusters, and the admixture proportions (Fig. S6) in our dataset using ~8.5 million SNPs. We explored the structure in our samples by running the admixture analyses with a different number of estimated ancestry clusters, ranging from 2 to 10 clusters. To ensure

convergence to the global maximum, the analysis was repeated 200 times, and the replicate with the best likelihood was chosen.

Admixture analysis using SNPs data

We compared the Zhokhov genome to a genotype diversity panel comprising multiple dog breeds using a model-based clustering algorithm. To allow for the inclusion of more sled dogs and reference individuals to complement full genomes, Illumina CanineHD 185,805 array genotyped data from 192 dogs was included solely in Figure S8 (Table S3). The Zhokhov genome was incorporated into the panel as diploid called genotypes. From each of the dog breeds included we randomly selected 10 individuals when more than 10 were available in the dataset, except for the sled dogs, for which we included all individuals available (Table S3). We kept sites with MAF≤0.01 and maximum missingness of 10%, and the final dataset consisted of 214 dog samples and 130,253 sites. ADMIXTURE was run on the final dataset assuming 2 to 12 clusters (K=2-12), and for each K value, we ran 50 replicates starting from different seeds (Fig. S7A). Additionally, we obtained cross-validation errors for the best replicate of each K (Fig. S7B).

Admixture graphs using TreeMix

We used TreeMix (*48*) to explore the broad phylogenetic context of the Zhokhov sample with respect to ancient and present-day wolf and dog genomes. From the whole-genome dataset, we selected samples representing the main ancestry groups (Table S1 - Selected): Eurasian grey wolves (n=19), Arctic grey wolves (n=6), Mexican grey wolves (n=2), Alaskan grey wolves (n=2), Pleistocene wolves (n=2), Asian dog breeds (n=6), European dog breeds (n=5), Siberian and Alaskan huskies (n=5), Alaskan Malamute dogs (n=1), Greenland village dogs (n=10), American pre-contact dogs (n=1), the Zhokhov dog, and Coyote (n=2), which was used as outgroup. For each sample and site, we randomly sampled one allele by building a majority count consensus sequence using ANGSD with a similar filtering approach as the one described in the MDS section. Allele frequencies were estimated on the different clusters as described in Table S1 (Selected). Additionally, we excluded invariant sites, transitions and sites with missing data. TreeMix was run on the final dataset which consisted of 766,082 sites assuming 0 to 7 migration edges (m=0-7). For each migration, we ran 100 replicates starting at different seed values and chose that with the highest likelihood for each value of *m* (Fig. S8).

CTVT

We analysed data from two publically available CTVT genomes (and their respective host genomes) (*49*), which possess a low level of host contamination and which were previously genotyped in Leathlobhair *et al*. (*3*). We used the same set of variable sites (~2.03M SNPs, including ~600K transversions) described in Leathlobhair *et al*. (*3*) for all following analyses including CTVT genomes. Briefly, this set was selected in order to minimize the effect of somatic mutations, host contamination, and to retain only regions of the dog reference genome that are diploid in cancer cells, as previously described in (*49*). We used plink v1.9 (*50*) to compute an Identity By State (IBS) matrix using all 2.03M SNPs. This matrix was used to build a neighbour joining tree (NJ) using the R package "ape" (*51*). The resulting tree shows that the CTVT genomes (C_79T and C_24T) cluster with the Port au Choix dog in clade basal to the Zhokhov and sled dogs (Fig. S9). We then computed outgroup f3-statistics as f3(CTVT, X; Andean Fox) using ADMIXTOOLS (*52*) where X is any other dog population to quantify the

amount of genetic drift shared between the CTVT other dogs using only transversions (Fig. S10). This analysis recapitulated the same pattern, indicating that the CTVT genomes are indeed closer to the Port au Choix genome. Reassuringly, the CTVT genomes did not cluster with their respective hosts (H_79T and H_24T in Fig. S10).

D-statistics

D-statistics as implemented in ADMIXTOOLS (*52*) were used to evaluate the shared ancestry and gene flow between the Zhokhov genome, modern and ancient dog and wolf genomes (Figs. S11 - S15). In order to incorporate genomes with heterogeneous depth of coverage, we randomly sampled alleles from a majority count consensus sequence at every position and for each sample. Reads with quality lower than 30, bases with quality lower than 20 and sites with coverage lower than 3 were discarded from the analysis. Additionally, transitions were removed from the final dataset to avoid incorporating errors derived from ancient DNA damage in the Zhokhov genome and other ancient samples used. A weighted block jackknife procedure over 1Mb blocks was used to assess the significance of the tests. Deviations from D=0 with a Z-score above or below 3.3 ($|Z|>3.3$) were presumed significant. For most D-statistics tests, we use the Andean fox (*Lycalopex culpaeus*) genome as an outgroup to avoid conflicting results caused by the frequent interbreeding between canids (*53*).

a) Zhokhov is more closely related to dogs than to wolves

To test if the Zhokhov genome form a clade with dogs to the exclusion of wolves, we computed a D-statistic of the form D(Zhokhov, H2; Croatian wolf, Andean fox), where H2 corresponds to dog genomes in the reference panel. In each case, we find support for the Zhokhov genome forming a clade with dogs to the exclusion of the Croatian wolf (Fig. S11). Conversely, we were able to reject the alternative hypothesis that the Zhokhov genome form a clade with the Croatian wolf or falls basal to wolves and dogs ($|Z|>40$).

b) Zhokhov dog falls basal to the Greenland sled dogs and other sled dogs

Admixture graphs obtained from TreeMix suggested Zhokhov dog lineage diverged from the ancestor of present-day Greenland sled dogs and other sled dogs (Alaskan malamute, Alaskan husky and Siberian husky). To further confirm its phylogenetic positioning within dogs, we computed a D-statistic of the form D(H1, H2; Zhokhov, Andean fox), where H1 and H2 represent all possible pairs of Greenland sled dogs and other sled dogs (Fig. S12). We found the Zhokhov dog to be symmetrically related to every pair of Greenland led dogs and pairs of other sled dogs ($|Z|\square 3.3$) however, when comparing Greenland sled dogs with other sled dogs, the ancient Zhokhov sample was found to be significantly closer to Greenland sled dogs ($|Z|>3.3$). Additionally testing of affinity of modern sled dogs to the Zhokhov dog, using D(Alaskan Husky 1, H2; Zhokhov, Andean fox), further supports more allele sharing between the Zhokhov dog and Greenland sled dogs, than to other sled dogs. Finally, using test D(Greenland sled dog, Siberian Husky 2; H3, Andean Fox), of all ancient dogs in the data, the Zhokhov dog shared most alleles with Greenland sled dogs.

c) Affiliation of Greenland sled dogs and other sled dogs, outside Zhokhov

TreeMix results suggested other sled dogs, particularly Siberian and Alaskan huskies, carry gene flow from other, possibly European, dog breeds (Fig. S8). We used D-statistics to test if this gene flow could explain the significant results obtained in the test. First, we computed a D-statistic of the form D(other sled dog, Greenland sled dog Aasiaat 1; H3, Andean fox), where H3 represent all dogs in the reference panel and Aasiaat 1 was used as a representative of the Greenland sled dogs, to identify the best candidate for the gene flow observed in other (non-Greenland) sled dogs (Fig. S13A). Next we used the form D(sled dog 1, sled dog 2; H3, Andean fox), where H3 represents 4 diverse dogs in the reference panel and sled dog 1 and sled dog 2 represent all possible pairs of sled dogs to identify the best candidate for the gene flow observed in other (non-Greenland) sled dogs (Fig. S13B). The statistics reflect the TreeMix results, finding other sled dogs, particularly Siberian and Alaskan huskies, carry gene flow from other dogs.

d) Affinities between the Pleistocene wolf, the Zhokhov dog and sled dogs

Previous studies have identified allele sharing between a Pleistocene Siberian wolf genome and dogs of especially Arctic context such as sled dogs (*5*), but also in extinct dogs from the Americas before European contact (*3*). We estimated a D-statistic of the form D(sled dogs, Boxer; Taymir/Yana, Andean fox) xy-plot (Fig. 1D and S14), and replicated these findings. However the signal was only significant when using the Yana wolf, likely due to fewer sites when using the lower coverage Taymir wolf genome. Similar to the previously documented Siberian Pleistocene wolf gene flow, we find that the Zhokhov dog shares significantly more alleles with Yana and Taimyr wolf, compared to the boxer dog.

e) Affinities between sled dogs and modern wolves

It is often heard that sled dogs are admixed with wolves and historical observations document such events in Arctic North America and Greenland (*54*, *55*). We tested this notion specifically using Greenland- and other sled dogs against a large panel of modern wolves, including specimens from the entire North American Arctic and Greenland (*53*, *56*) (Fig. S15). To estimate whether Greenland dogs admixed with local wolves since they diverged from their common ancestor with the Eurasian wolves, we computed a D-statistic of the form D(Greenland sled dogs, Portuguese wolf; Arctic wolves, Andean fox), for all Greenland dogs and all Arctic wolves. We found no significant admixture between any of the Greenland dogs and any wolves.

Population size estimates in Greenland sled dogs

We estimated the effective population size of the Greenlandic sled dogs, using the diffusion approximation model for site frequency spectrum (SFS) implemented in the package *moments*. First, we estimated the site frequency spectrum for the Greenlandic sled dogs, using the 11 samples across Greenland. We estimated the 1-D SFS for these sled dogs using *realSFS*, a utility implemented in *ANGSD* v0.921, which employs genotype likelihoods to incorporate uncertainty introduced by low and different coverage sequencing among the 11 samples. Using this SFS, we estimated parameters of a simple demographic model that allows for two population size changes (Fig. S16). The first one at 89,825 years ago, outside the relevant time frame of this study, the second 864 years ago corresponding with the Inuit/Thule culture's introduction of this dog lineage into Greenland. Demographic parameters were estimated under a maximum likelihood model for the SFS under the given class of demographic models.

Heterozygosity

Heterozygosity is often used as a proxy measure for effective population size. Here, we computed heterozygosity for the different whole-genome sequenced samples included in the study, using ANGSD v0.921 and realSFS (a utility tool from ANGSD). To compute heterozygosity, we estimated the single sample site frequency spectrum (SFS) for each sample. Folded SFS was estimated using genotype likelihoods computed using ANGSD. Heterozygosity was computed as the proportion of sites with a minor allele count of 1, which is effectively heterozygosity since the SFS is computed for a single sample. Finally, we obtained bootstrap variance of heterozygosity by obtaining 100 bootstrap estimates of the SFS. Estimated heterozygosity for the different samples (Fig S17A) and populations (Fig. S17B) included in this study showed lowest heterozygosity for Greenland sled dogs, among all non-breed dogs. The excess heterozygosity of other sled dogs, compared to the Greenland sled dogs, can be explained by genetic contributions from European dogs. The low heterozygosity of the Greenland sled dogs is consistent with the population size reduction shown in the demographic inferences (see above section).

Genetic adaptations in sled dogs

Population branch statistic

To detect signals of positive selection, we used the population branch statistic (PBS) (*10*), which identifies alleles that have experienced strong changes in frequency in one population (sled dogs) relative to two reference populations (a sister population, such as other dogs [non-sled dogs], and an outgroup, such as wolves). Thus, PBS identifies genomic regions highly differentiated in the sled dog branch - a signal that can be suggestive of positive selection. PBS has previously proven useful for identifying population specific selection in humans (*10*, *25*). We computed F_{ST} between the following three populations: sled dogs, other dogs and wolves (Table S1 - Selected). F_{ST} was computed using the Hudson estimator (*57*) in windows of 100 kilo-base pairs with a 20 kilo-base pairs slide (Fig 3A) and 25 kilo-base pairs with a 5 kilo-base pair slide (Fig S19). Since we were interested in discovering putative targets of positive selection, we used the alignments to the dog reference genome (See Quality control and alignment), taking advantage of its longer continuity and better gene annotation than the wolf assembly. The three F_{ST} measures were then combined in all approaches to obtain the PBS as described in (*10*), taking as the focal population sled dogs (Fig 3A) and other dogs (Fig S18). Genes overlapping with extreme outlier regions are listed in Table S4-S5.

XP-CLR

XP-CLR (*58*) was computed across the autosomes using xpclr 1.0 (https://reich.hms.harvard.edu/software) with parameters "*-w1 0.0075 200 500 chromosome -p0 0.95*". The two populations compared were sled dogs (n=17) and other dogs (n=61) (Table S1), as defined in our PBS scan. XP-CLR around the autosomal genomic regions above the 99.95th percentile of PBS can be found in Figure S23.

Functional enrichment in high PBS regions

We computed enrichment of gene ontology (GO) terms in the regions of the genome that displayed high PBS values when using the Greenland sled dogs as the focal population, to identify biological processes or molecular functions that genes in this region represent. To compute enrichment of GO terms in this region, we used the interval enrichment program, inrich (*59*), with dog specific annotations from the GO database amigo (http://amigo.geneontology.org/) and gene annotations from ensembl (. Further, we computed the enrichment using two different PBS thresholds - 0.451 and 0.328 (corresponding to the 99.95th and 99.5th percentile of the empirical distribution) - that resulted in 14 and 113 non-overlapping intervals respectively. The enriched GO terms with empirical p-value < 0.05 are shown in Table S6.

Supplementary Figures

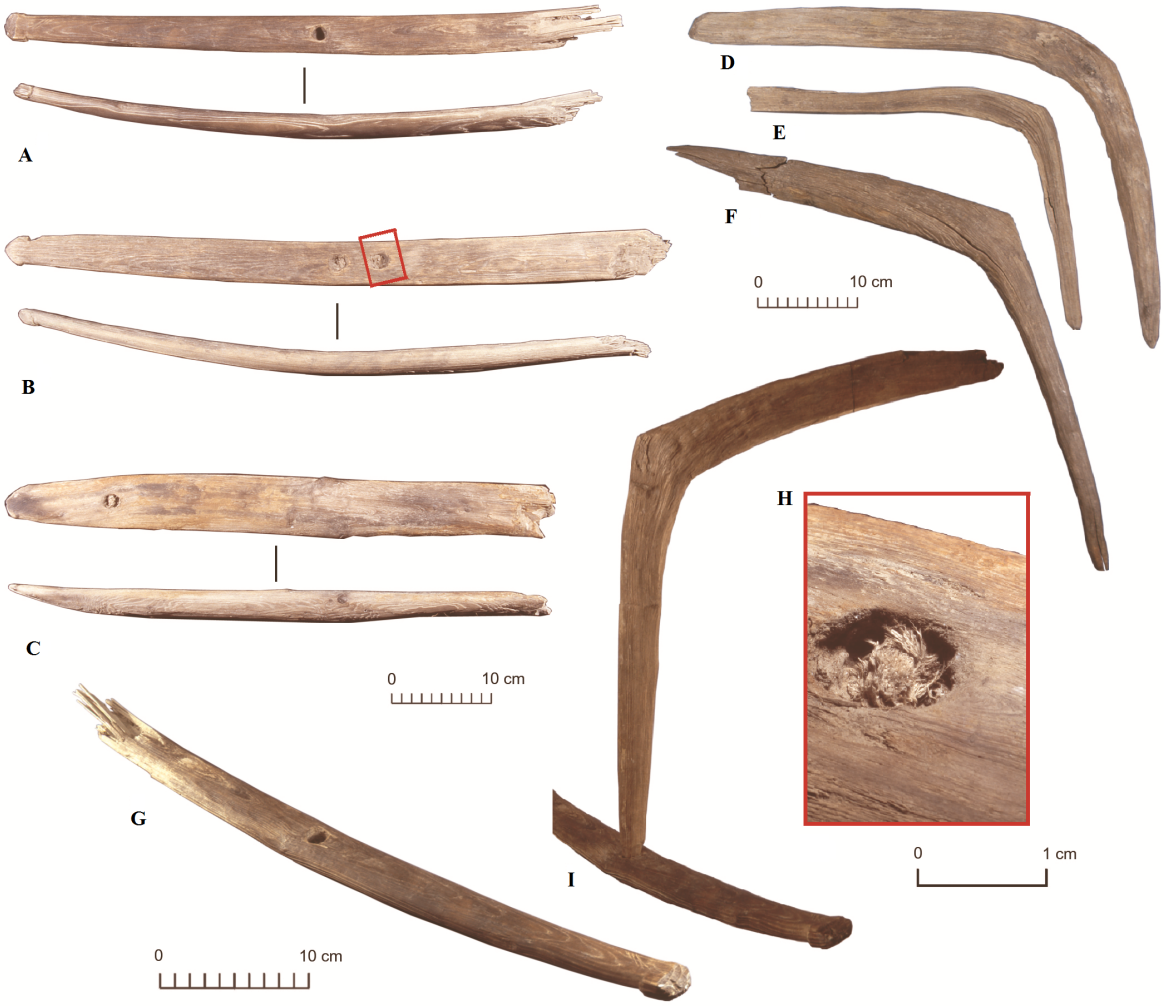
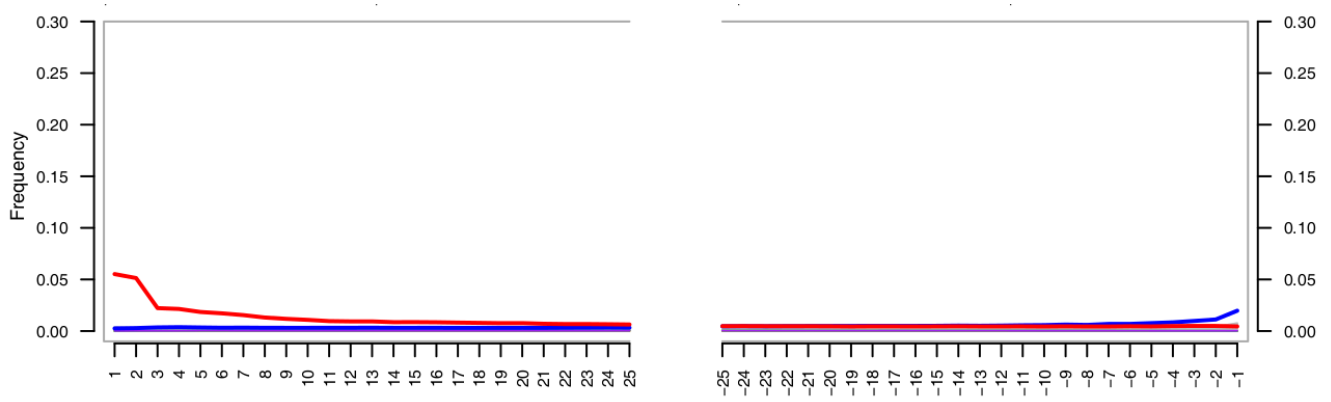


Fig. S1. The Zhokhov site sledge remains. (A-I) – sled runner fragments; **(D, E, F)** – upright **(F)** is combined with a sled runner **(G)** in **(H)** – hole with a piece of rope made of animal hair, as an example of using these parts together **(I)**; Parts indicated a large sled, making it unlikely that it was pulled by humans, but rather by dogs.

Fig. S2. The Zhokhov dog. Zhokhov dog (Zh-03-97 or CGG6) mandible and size in cm.

Fig. S3. The Yana wolf. Yana wolf (Y06-NP-18994 or CGG23) mandible and size in cm.

A



B

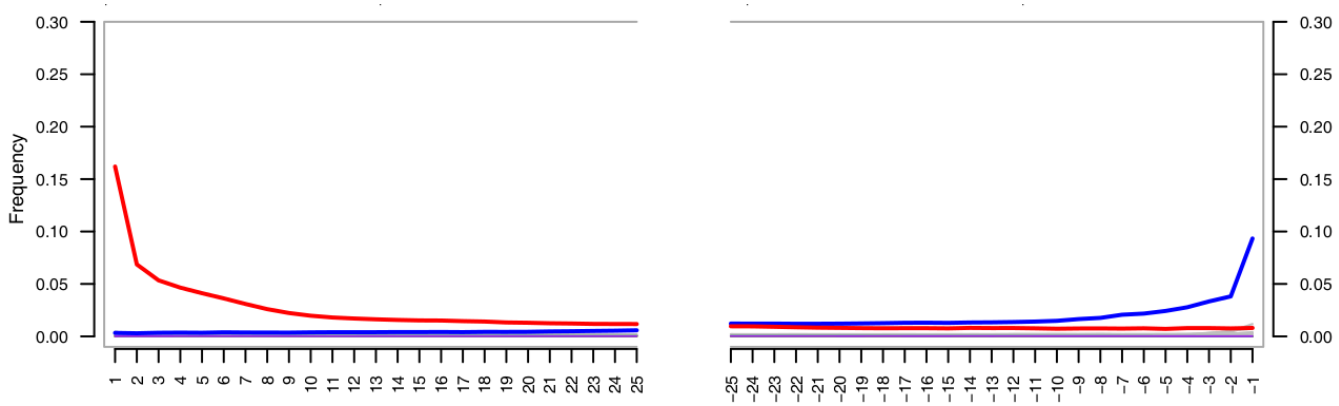


Fig. S4. Authenticity of ancient DNA (aDNA) sequencing data. Panels (**A**) and (**B**) show the different substitution rates at the 5' and 3' ends of reads in the Zhokhov dog and the Yana wolf sample respectively. In each panel, the plot on the left shows the substitution rates at the 5' end of reads, whereas the plot on the right shows the substitution rates at the 3' end of reads. The red line shows the C-T substitution rate, which is the most abundant substitution at the 5' end of reads due to ancient DNA damage. Similarly, the blue line shows the G-A substitution, which is the most abundant substitution at the 3' end of reads in both samples.

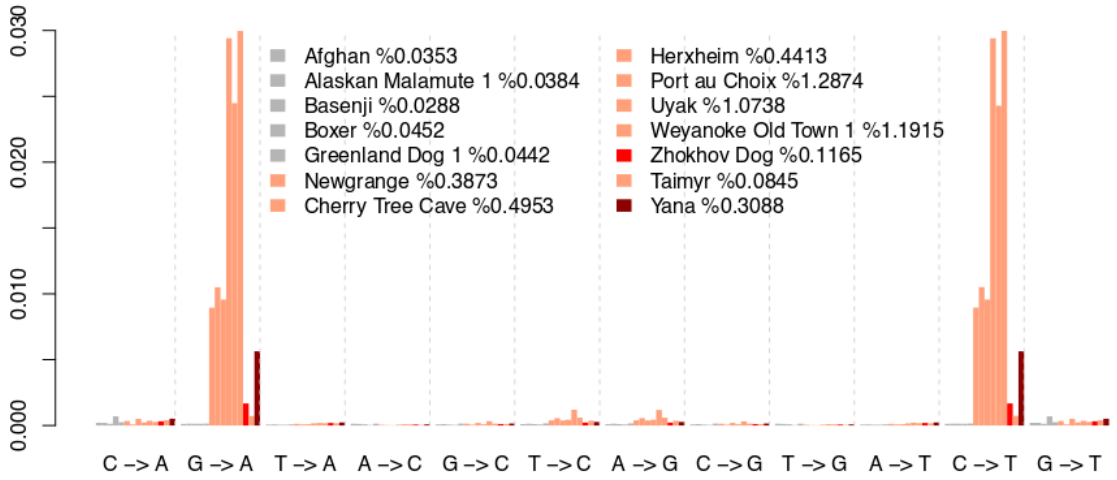


Fig. S5. Authenticity of ancient DNA (aDNA) sequencing data. Type specific error rates. Type specific error rates estimated using ANGSD for the ancient samples used in this study. Grey bars correspond to the error rates estimated in four modern dog genomes processed using the same pipeline as the ancient data and shown as comparison. Orange bars correspond to reference ancient samples used in this study, red and dark red correspond to the error rates estimated for the Zhokhov and Yana genomes sequenced in this study.

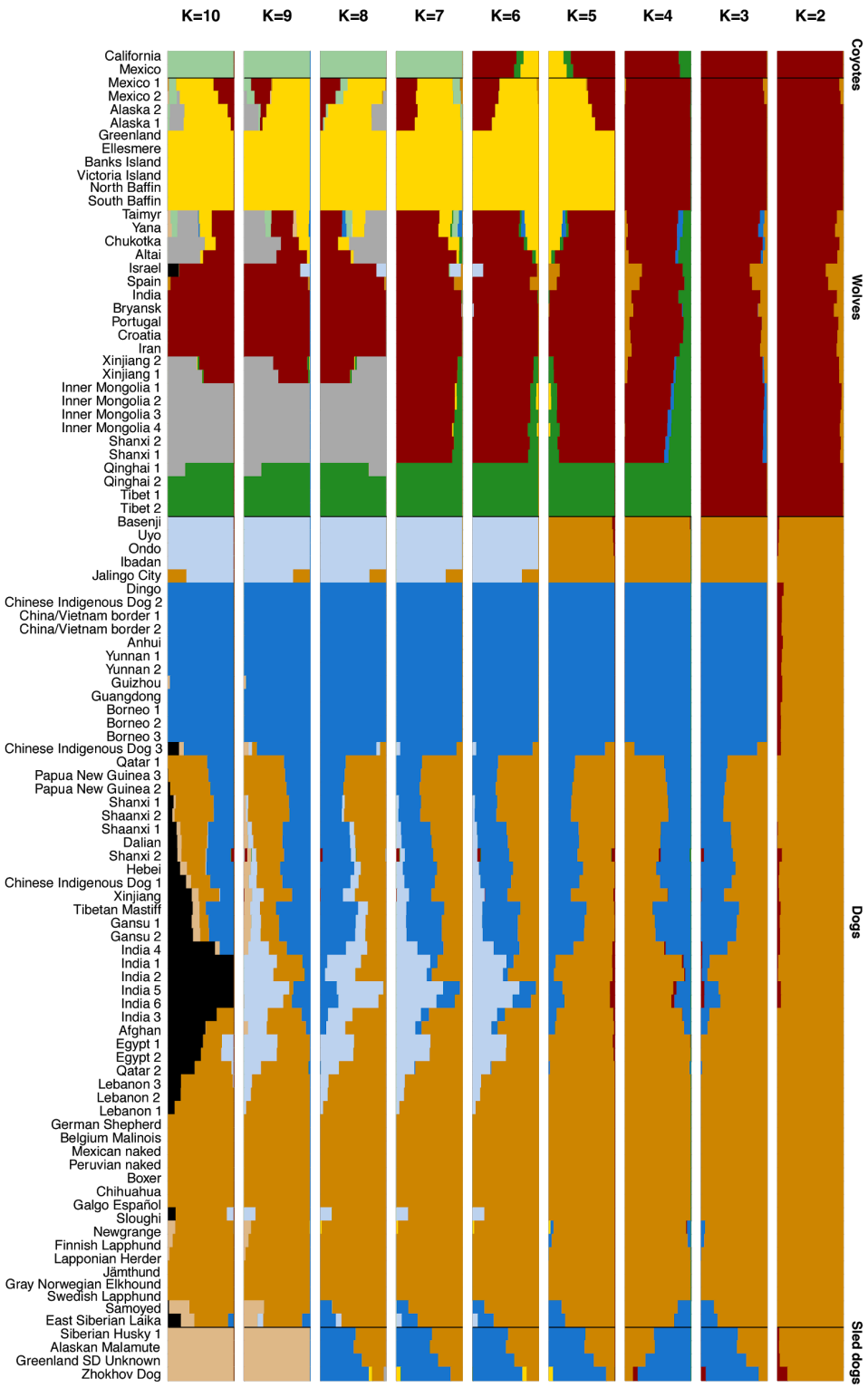


Fig. S6. Extended admixture results from whole-genome data. NGSadmix clustering results obtained for a panel comprising whole-genome data. Individual colors illustrate the inferred admixture components, Ks values indicated at the top. The sample name is given at the left and the relevant clusters are detailed at the right side of the figure.

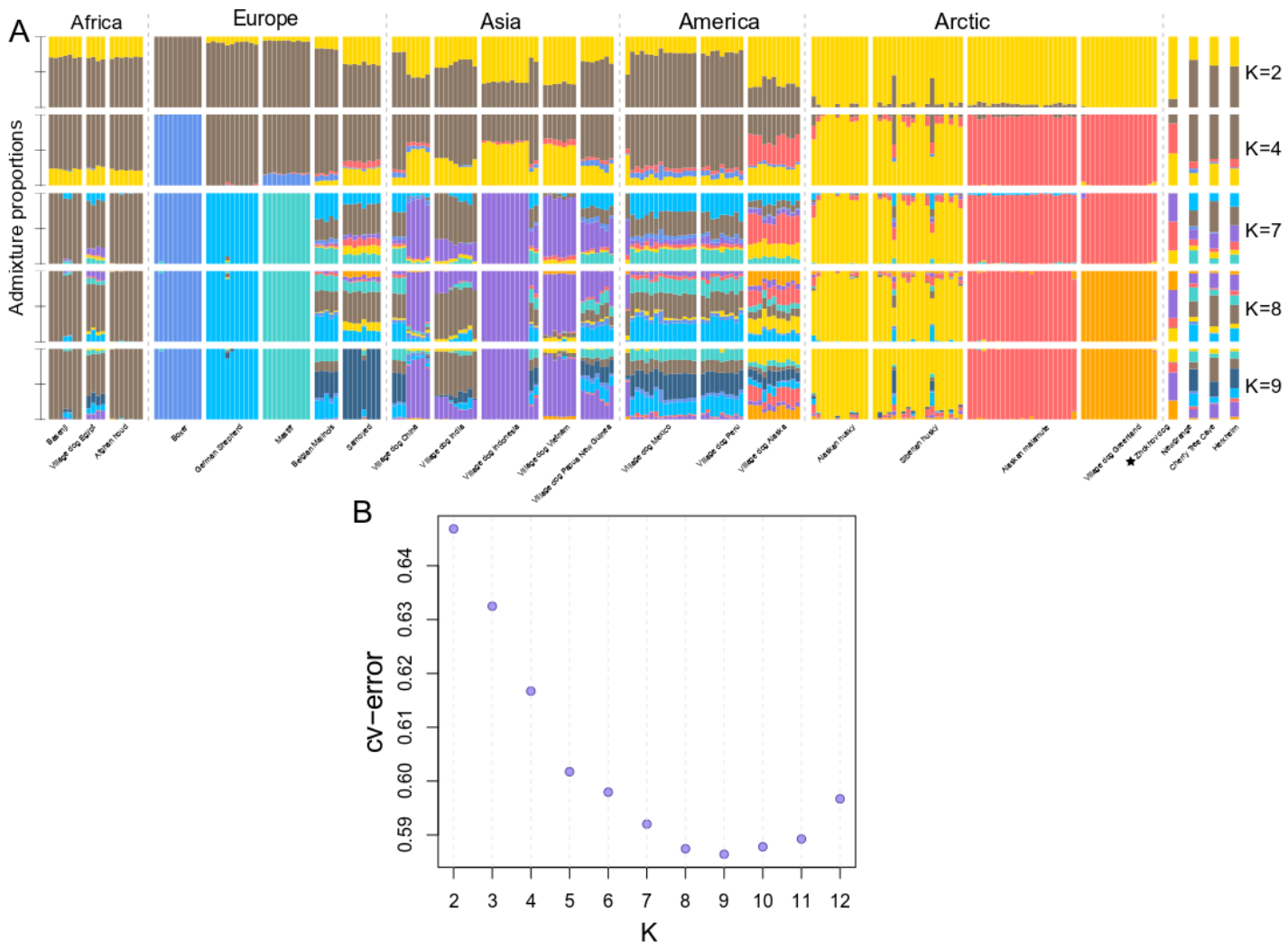


Fig. S7. Extended admixture results from called genotypes of genomes and genome wide SNP data. (**A**) Clustering results obtained using ADMIXTURE and a genotype panel comprising world-wide dog breeds. The Zhokhov sample was incorporated into the panel as called genotypes. Transitions, sites with >20% missing data and MAF<0.05 were excluded from the analysis leaving a total of 130,253 sites. ADMIXTURE was run assuming 2 to 9 clusters/populations (*K*=2-9). For each *K*, individual bars represent different samples and the colors represent the proportions of each of the inferred components. (**B**) Cross-validation errors obtained for each value of K.

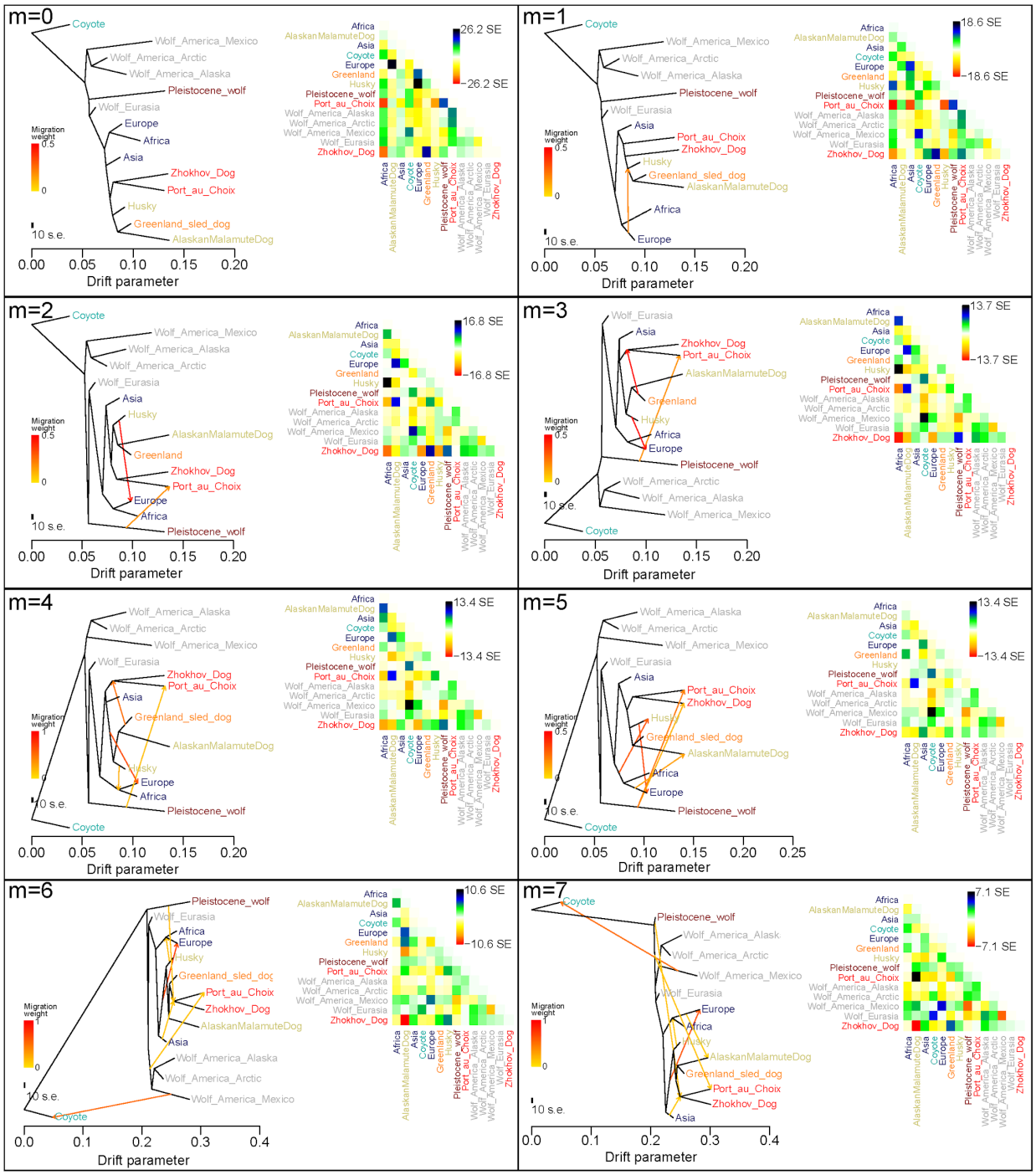


Fig. S8. Phylogenetic placement of Zhokhov sample using TreeMix admixture graphs. TreeMix relationship of major clusters of diversity in the dataset. Admixture graphs computed using TreeMix on a dataset consisting of 66 individuals merged into 15 groups, representing the major groups of wolves (American, Eurasian and Pleistocene wolves) and dogs, and enriched for sled dogs dogs. For each individual a random allele was chosen at each site; transitions, non-polymorphic sites and sites with missing data were excluded from the analysis (final dataset of 766,082 sites). We fitted from 0 to 7 migration edges. In the right of each subpanel; inferred tree

with admixture edges represented as arrow and colors indicating the fraction of admixture. In the left of each subpanel, a heatmap indicating the residuals obtained from the fitted graph with colors indicating the standard errors of each node.

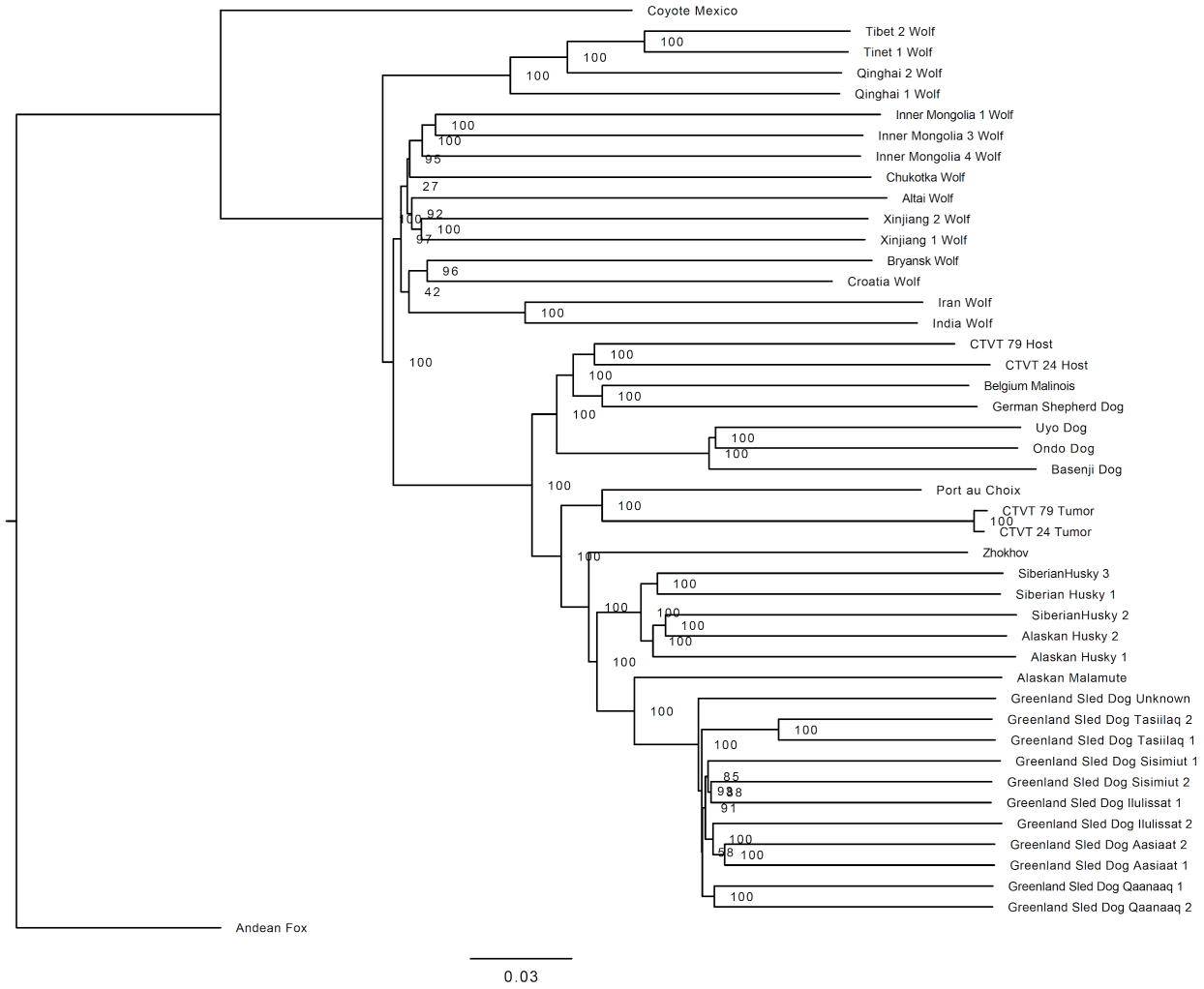


Fig. S9. CTVT NJ tree. A neighbour joining (NJ) tree illustrating cladistic relationship between CTVT genomes, wolves and dogs. The NJ tree is based on Identity By State and bootstrap values supporting phylogenetic clusters. The tree place the CTVT genomes closer to the Port au Choix dog than to Zhokhov and sled dogs. Further the tumor hosts are correctly placed near European dogs.

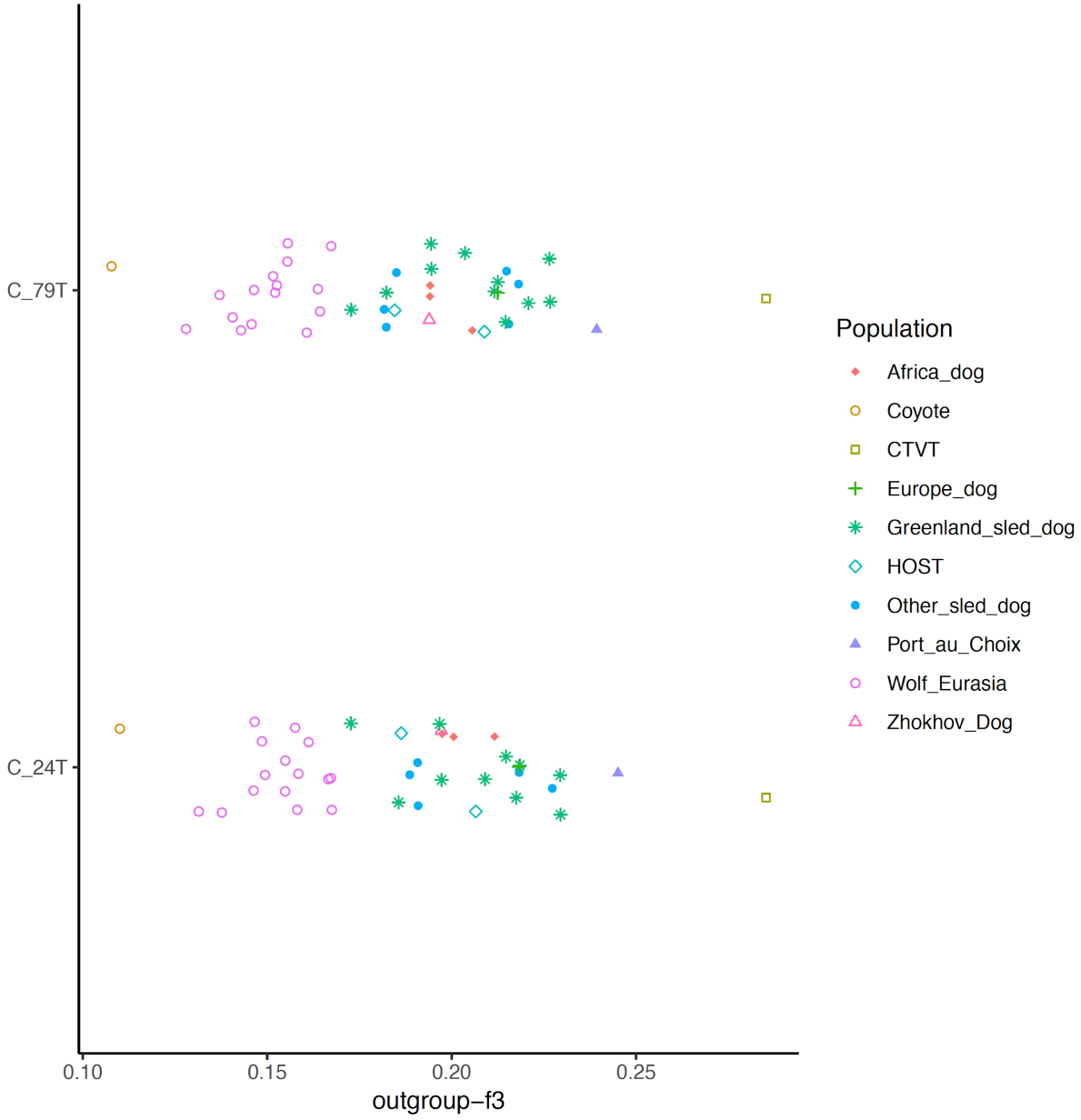


Fig. S10. CTVT f3-statistics. f3-statistics testing the relationship between CTVT genomes, wolves and dogs. Shared genetic drift measured by f3(Outgroup; Y, X) where X is either Tumor C_79T or C_24T and X is a selection of references (same individuals as in Fig. S10) shown in panel to the right.

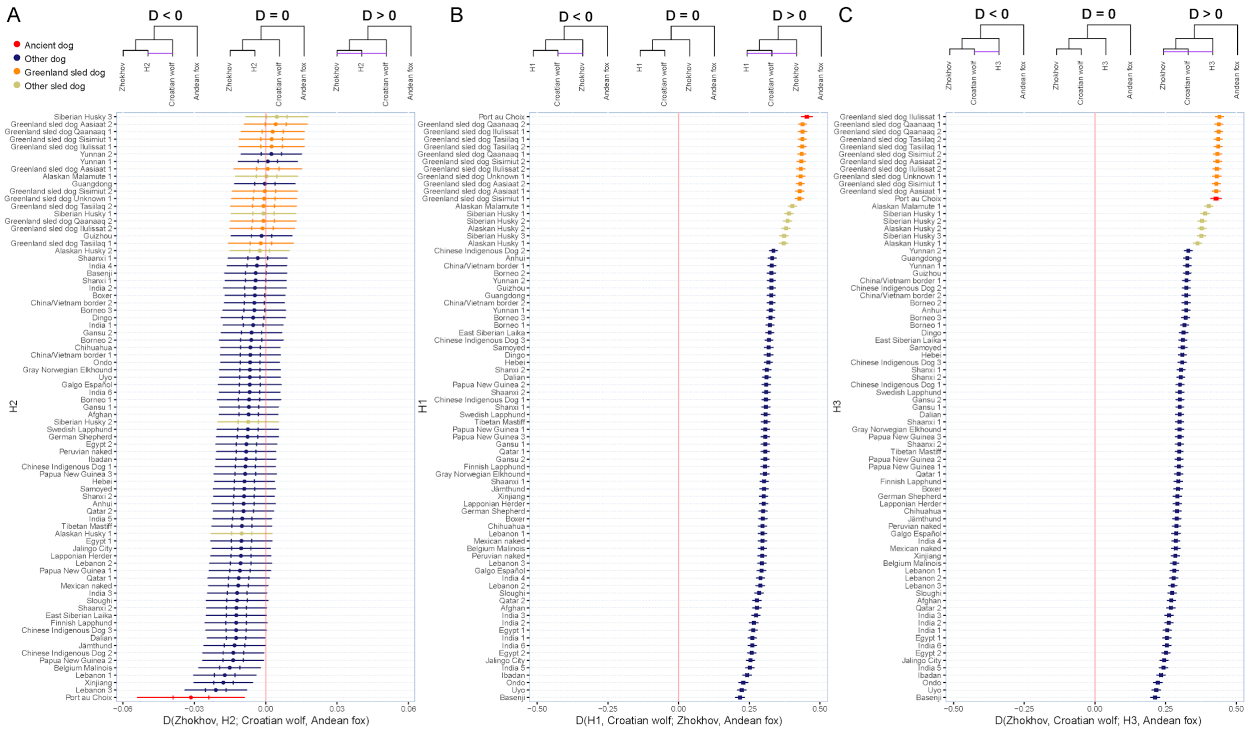


Fig. S11. D-statistics showing Zhokhov genome is a dog. D-statistics testing the relationships between the Zhokhov genome, dogs and wolves. (A) Test showing the Zhokhov genome forms a clade with all dogs to the exclusion of the Croatian wolf ($-3.33<Z<3.33$). On the contrary, tests where (B) the Zhokhov genome is an outgroup to pairs of dogs and the Croatian wolf, or (C) the Zhokhov dog forms a clade with the Croatian wolf to the exclusion of all dogs were all rejected ($Z>3.33$). D-statistics were computed using whole-genome data and a random allele for each sample as described in the methods section. Transitions were not included in the analysis. Individual points represent the *D* value obtained from each test. Horizontal bars show 1 (first vertical mark) and ~3.3 standard errors. Tree topologies at the top of each panel indicate the null ($D=0$) and alternative hypotheses ($D>0$ and $D<0$) tested.

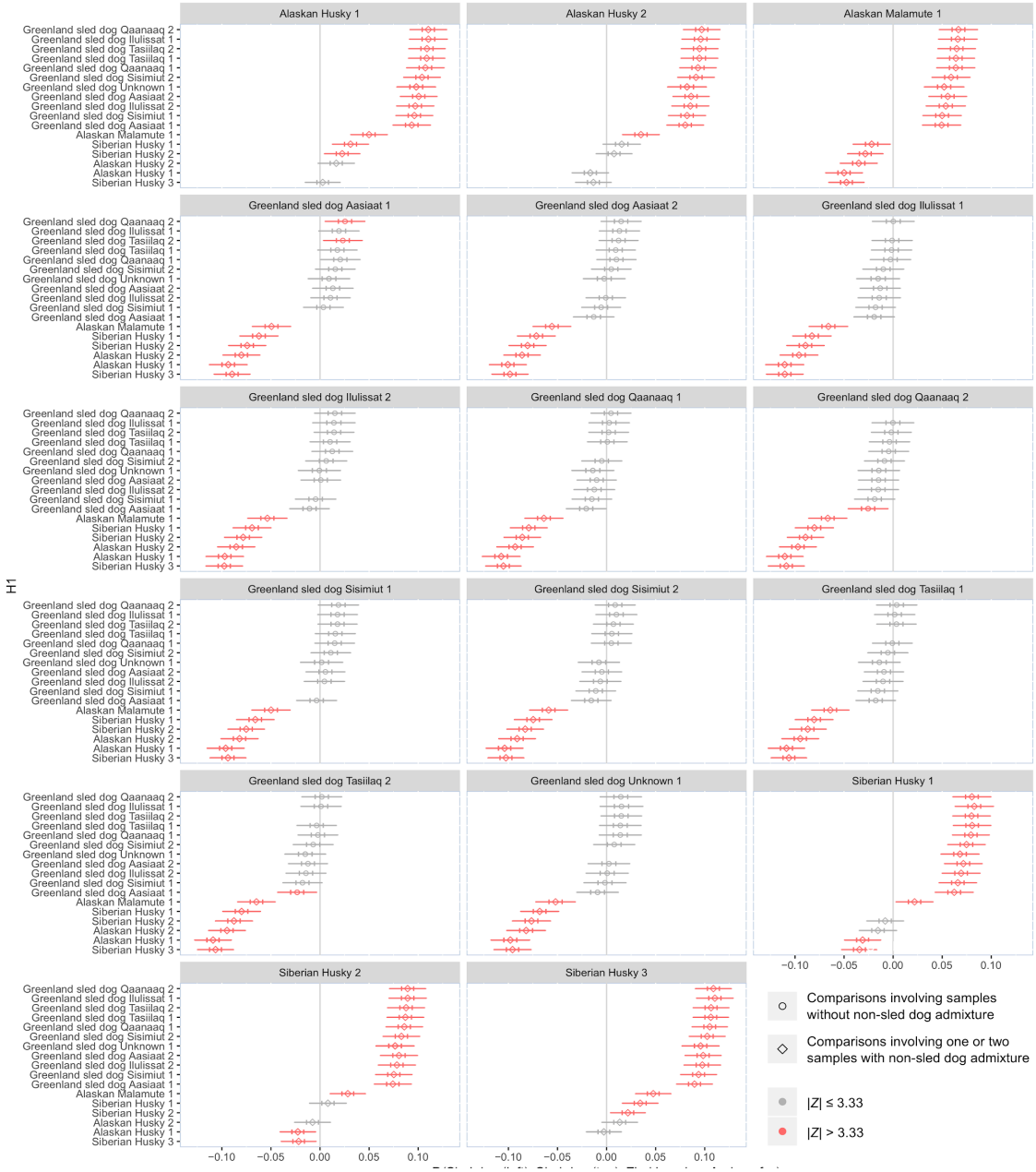


Fig. S12. D-statistics showing Zhokhov genome falls basal to sled dogs. D-statistics supporting the position of Zhokhov dog as an outgroup to sled dogs. We computed D-statistic tests of the form D(H1, H2; Zhokhov dog, Andean fox), where H1 and H2 represent all possible combinations of sled dog genomes available. Values of D are indicated as individual points. Horizontal bars represent 1 (first mark) or ~3.3 standard errors. Tests involving pairs of Greenland sled dogs yielded values of D consistent with the Zhokhov dog being symmetrically related. Most tests involving pairs of Other (non-Greenland) sled dogs resulted in values of D that suggested that either a) the Other (non-Greenland) sled dogs carried admixture from some sample outside the sled dogs or b) the tree topology suggested in the test is incorrect. In Fig. S13, we show that the a) is the most likely explanation. Tests that yielded significant scores ($|Z|>3.33$) are shown in red.

A


Alaskan Husky 1
Alaskan Husky 2
Alaskan Malamute 1
Siberian Husky 1
Siberian Husky 2
Siberian Husky 3
H3
Ondo
Afghan
Lebanon 3
Uyo
India 1
Lebanon 1
Egypt 2
Basenji
Ibadan
Jalingo City
Galgo Español
Boxer
India 2
Lebanon 2
German Shepherd
India 6
India 3
India 5
Slought
Egypt 1
Qatar 2
Chihuahua
Peruvian naked
Xinjiang
Mexican naked
Jämthund
Finnish Lapphund
East Siberian Laika
Gray Norwegian Elkhound
Shanxi 2
Samoyed
India 4
Gansu 1
Swedish Lapphund
Lapponian Herder
Gansu 2
Papua New Guinea 3
Hebei
Shaanxi 1
Qatar 1
Papua New Guinea 1
Papua New Guinea 2
Chinese Indigenous Dog 1
Shaanxi 2
Tibetan Mastiff
Shanxi 1
Dalian
China/Vietnam border 2
Guangdong
China/Vietnam border 1
Borneo 3
Yunnan 1
Borneo 2
Yunnan 2
Guizhou
Chinese Indigenous Dog 3
Chinese Indigenous Dog 2
Borneo 1
Anhui
Dingo
−0.04
0.00
0.04
D(Other sled dog (top), Greenland sled dog Aasiaat 1; Dog (left), Andean fox)
|Z| < 3.33
|Z| ≥ 3.33

B

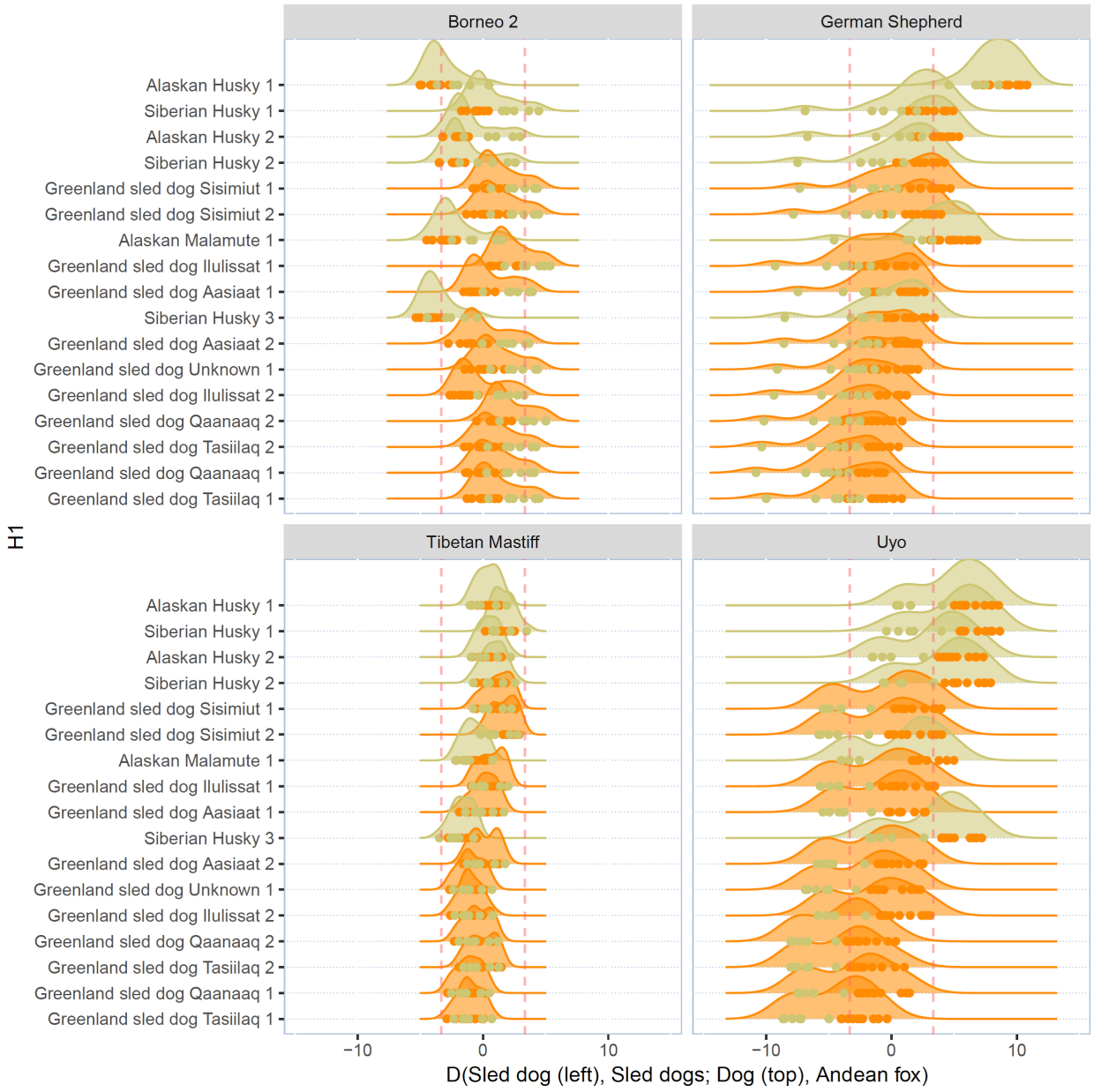


Fig. S13. Affiliation of Greenland sled dog and other sled dogs, outside Zhokhov. D-statistics testing for gene flow from non-sled dog breeds into the sled dogs. **A.** We computed a D-statistic of the form D(Other sled dogs, Greenland sled dog Aasiaat 1; Dogs, Andean fox) in order to test whether Other (non-Greenland) sled dogs carry ancestry from other dogs breeds when compared to Greenland sled dogs. D-statistics were estimated using whole-genome data and a random allele for each sample as described in the methods section. Transitions were not included. Individual points represent the *D* value obtained from each test. Horizontal bars show 1 (first vertical mark) and ~3.3 standard errors. Tests indicating significant gene flow between samples in the ingroup and H3 ($|Z|>3.33$) are shown in red. **B.** D-statistic tests showing that sled

dogs resulting in significant deviations from the test D(sled dog, sled dog; Zhokhov dog, Andean fox) (Figure 2A), also resulted in significant gene flow from other dogs. Individual density distributions correspond to the z-scores obtained from all tests of the form D(H1 (left), all possible SDs; admixing dog (top), Andean fox), where the admixing dog is represented by four dogs that yielded significant scores ($|Z|\geq 3.3$) for the test in Figure S13. Points represent the Z-scores obtained from each test. Colors indicate whether the dog in H2 is a Greenlandic (orange) or another sled dog (yellow). Dotted lines show the significance threshold $|Z|\geq 3.3$. In brief, tests involving Greenland sled dogs do not show significant gene flow from any of the admixing dogs. In contrast, tests involving other sled dogs yielded significant deviations from D=0 ($|Z|\geq 3.3$).

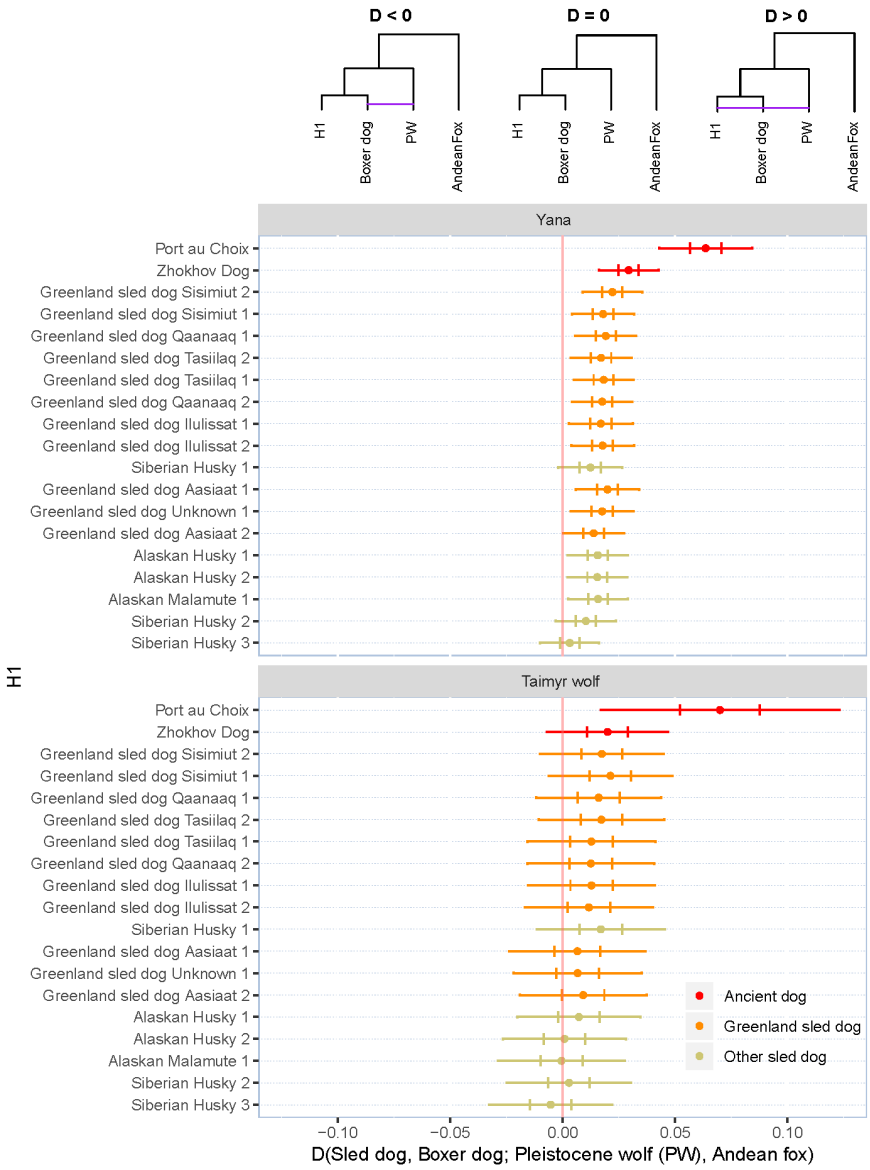


Fig. S14. Pleistocene wolves D-statistics. D-statistics showing allele sharing between the ancient Taimyr and Yana wolves and sled dogs. D-statistics showing there is excess allele sharing between both Pleistocene wolves (Yana and) most sled dogs, including the Zhokhov and American pre-contact dogs, when compared to the modern grey wolf. Note that, even though not all results yield significant Z-scores ($|Z|>3.33$) when using Taimyr, potentially due to the limited amount of sites resulting from the low coverage data, both Taimyr and Yana show the same pattern. D-statistics were estimated using whole-genome data and a random allele for each sample (see methods section). Transitions were excluded. Individual points represent the *D* value obtained from each test. Horizontal bars show 1 (first vertical mark) and ~3.3 standard errors.

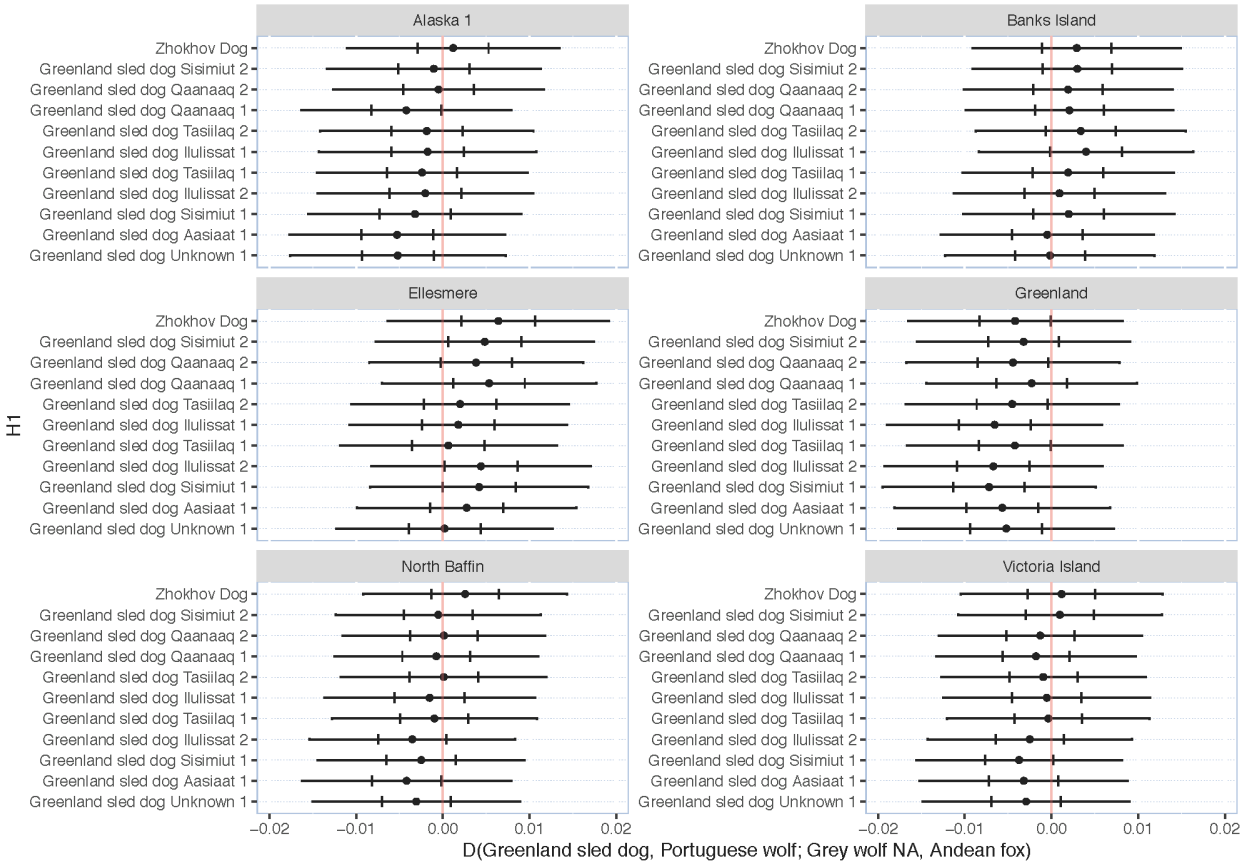


Fig. S15. Modern wolves D-statistics. D-statistics testing for gene-flow between Greenland sled dogs and a wolf from: Greenland, Victoria Island, Baffin Island, Alaska and Siberia. Individual samples are given on the left (H1) or in the top (H3) of each panel. D-statistics were estimated using whole-genome data and a random allele for each sample (see methods section). Transitions sites were excluded. Points indicate the *D* value obtained from the test. Horizontal bars show 1 (longer line) and ~3.3 (shorter line) standard errors. We did not find significant gene flow between any of the Greenland sled dogs and Arctic wolves tested.

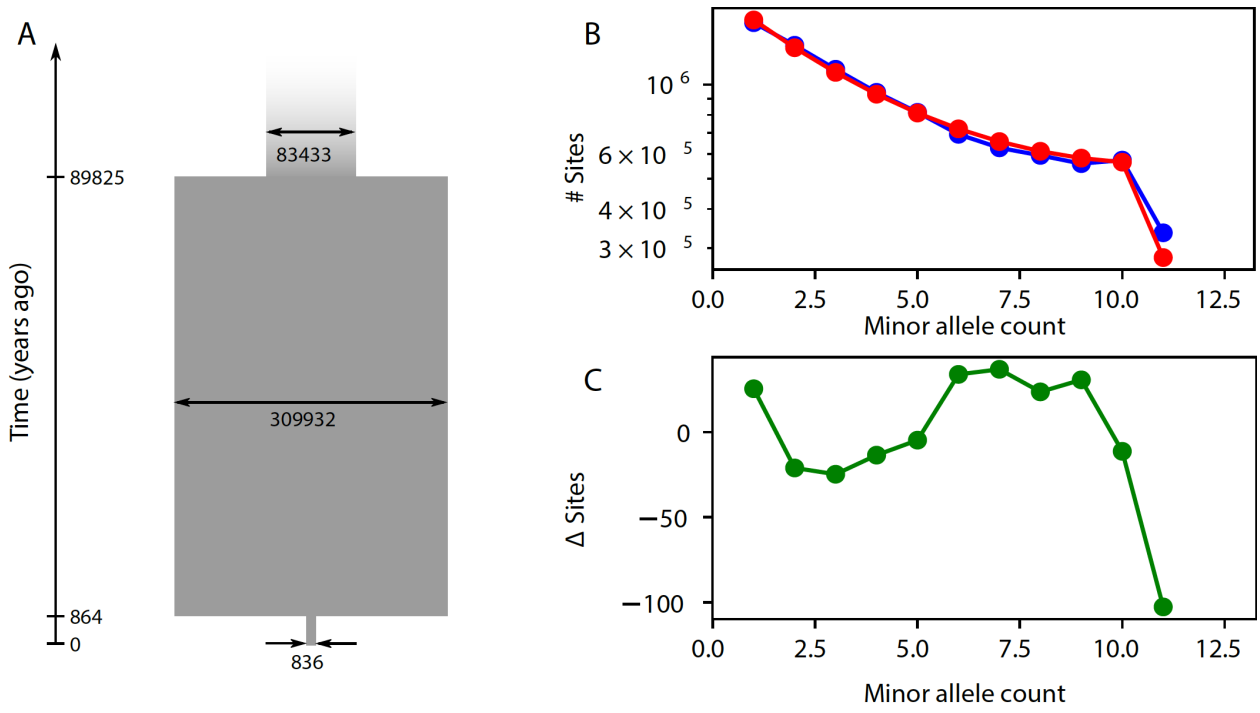
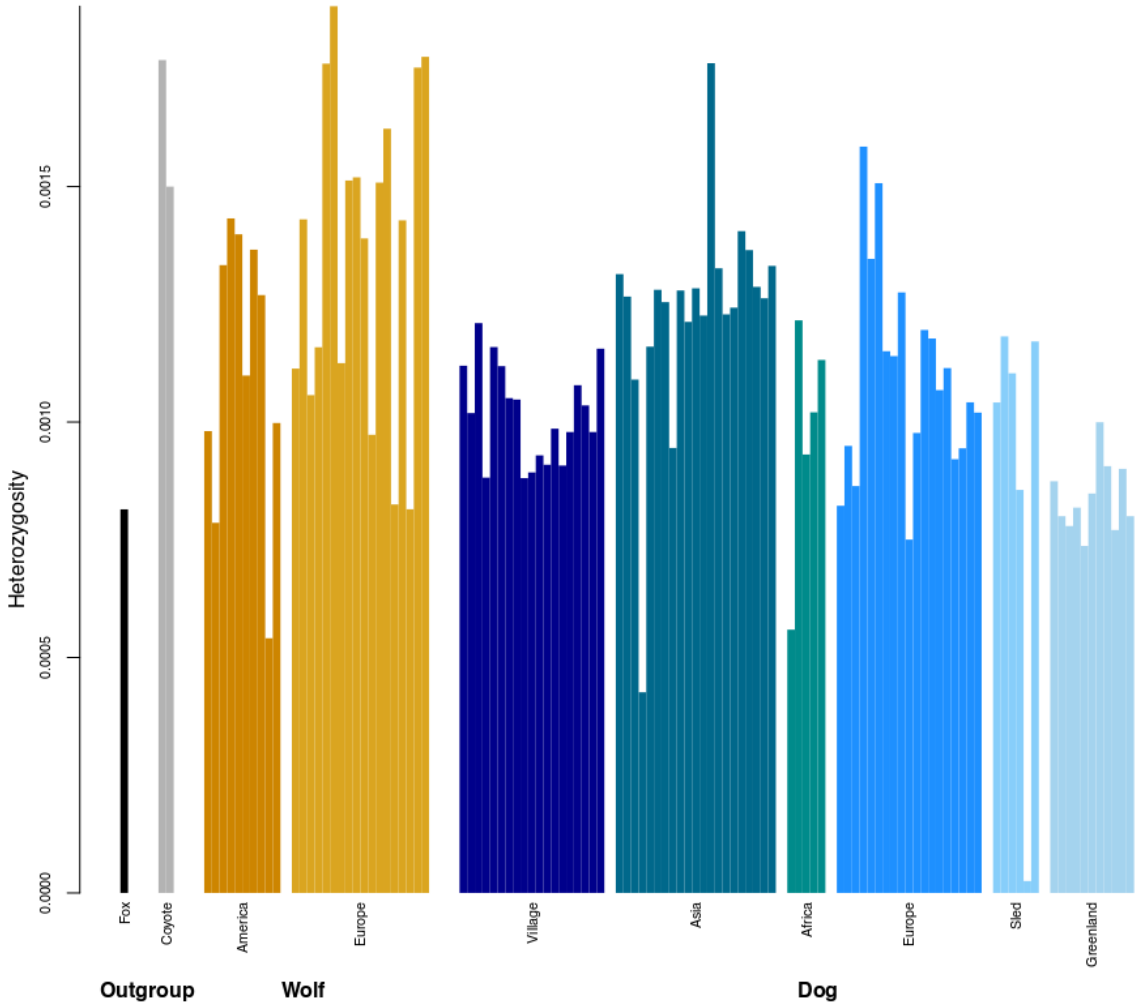


Fig. S16. Effective population size (N_e) of Greenland sled dogs through time. The demography of the Greenland sled dogs estimated using the diffusion approximation, as implemented in the software package *moments*. **A.** the demographic history of the Greenland sled dogs, with population size indicated by the width of the bar, and time on the y-axis, in years. **B**. The estimated (blue line) and observed (red line) folded site frequency spectra (SFS). **C.** Difference between the observed and estimated SFS. The y-axis shows the deviation of the estimated SFS from the observed SFS (in number of sites).

A


Heterozygosity
0.0000
0.0005
0.0010
0.0015
Fox
Coyote
America
Europe
Village
Asia
Africa
Europe
Sled
Greenland
Outgroup
Wolf
Dog

B

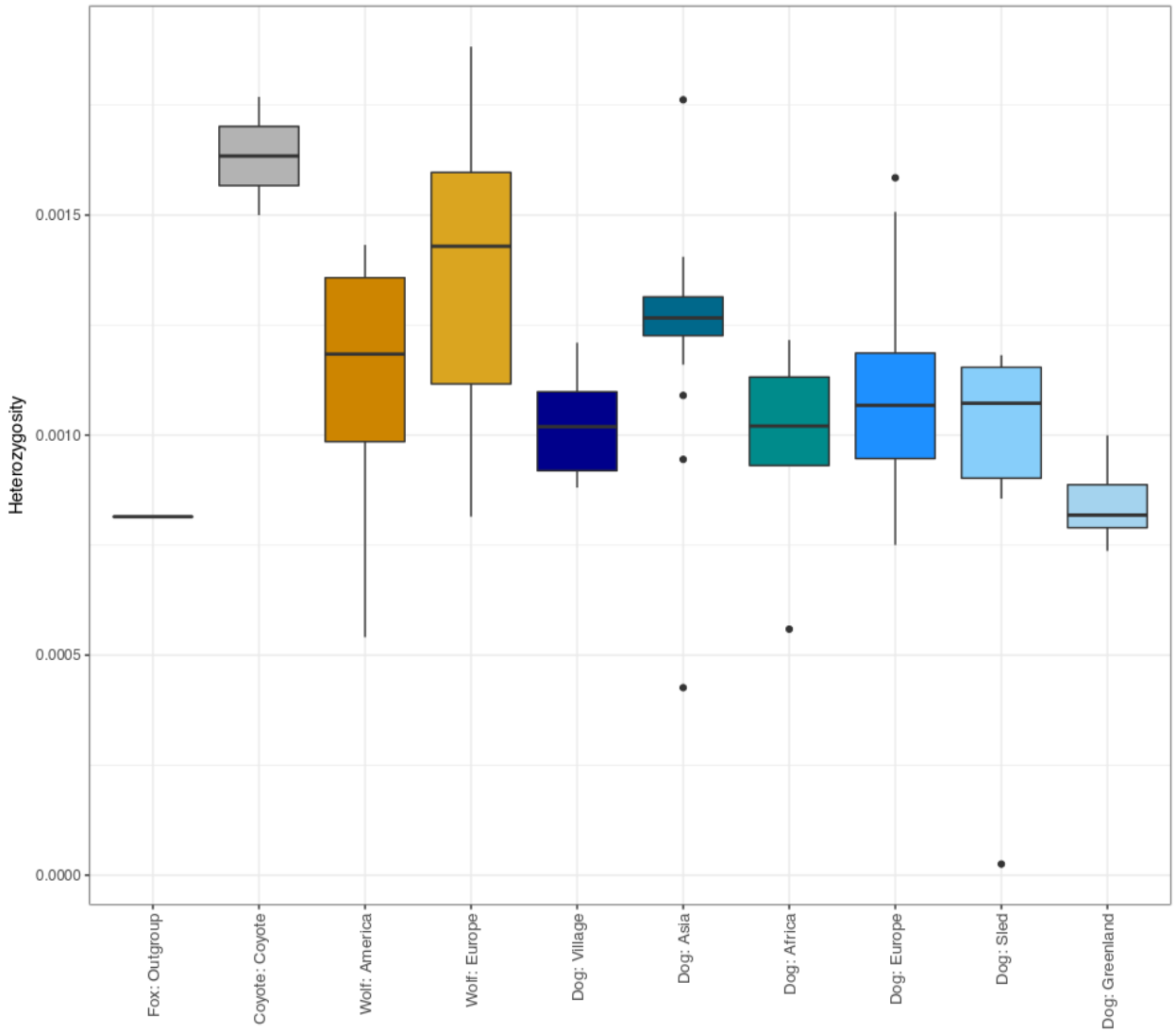


Fig. S17. Heterozygosity estimates. Heterozygosity estimated under a genotype likelihood framework. (**A**) Heterozygosity for each sample is represented by a single bar, and the samples are grouped by population labels. Error bars are not displayed. (**B**) Population heterozygosities plotted using box plots. The median heterozygosity is marked by the horizontal line, while the box marks the first and third quartiles. The whiskers extend from the end of the box 1.5 times the interquartile range, or to the end of the range, whichever is closer. Finally, observations beyond the whiskers are considered outliers. Overall, Greenland sled dogs have the among the lowest diversity in dogs, with the exception of a few other dogs - a Dingo, a Basenji and a Siberian Husky show lower heterozygosity.

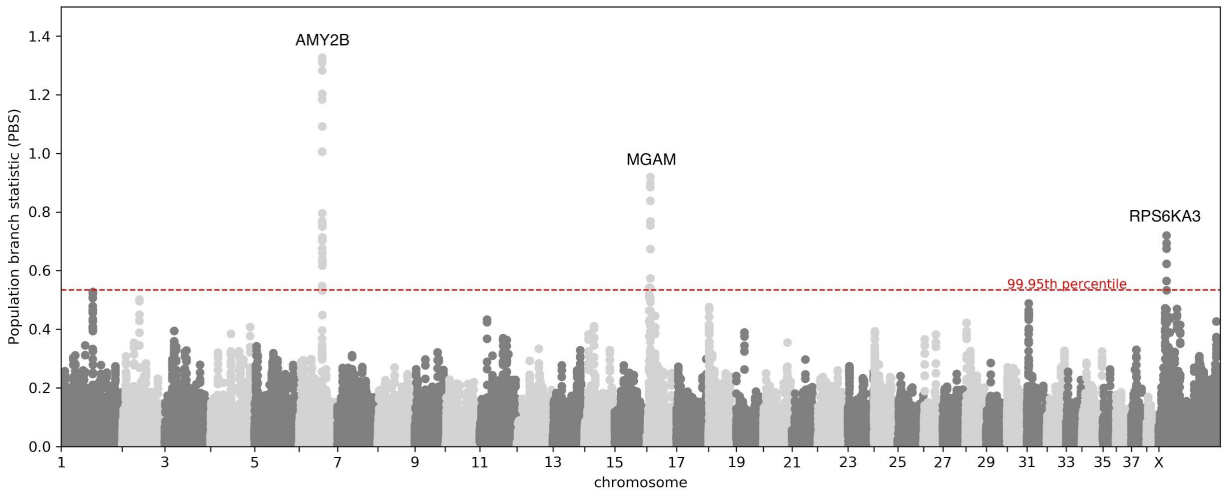


Fig. S18. Manhattan plot of PBS with other dogs as the focal population. PBS values in windows of 100 kilo-base pairs using a 20 kilo-base pairs slide. The 99.95th percentile of the empirical distribution is shown as a red dashed horizontal line. Names of genes with highest overlap associated with the peaks are shown. We note other genes not displayed in the figure overlap such regions: chr6: RNPC3, AMY2B, chr16: MGAM, TAS2R38, CLEC5A, COR9A7 and chrX: RPS6KA3, MAP7D2, EIF1AX. The full list can also be found in Table S5.

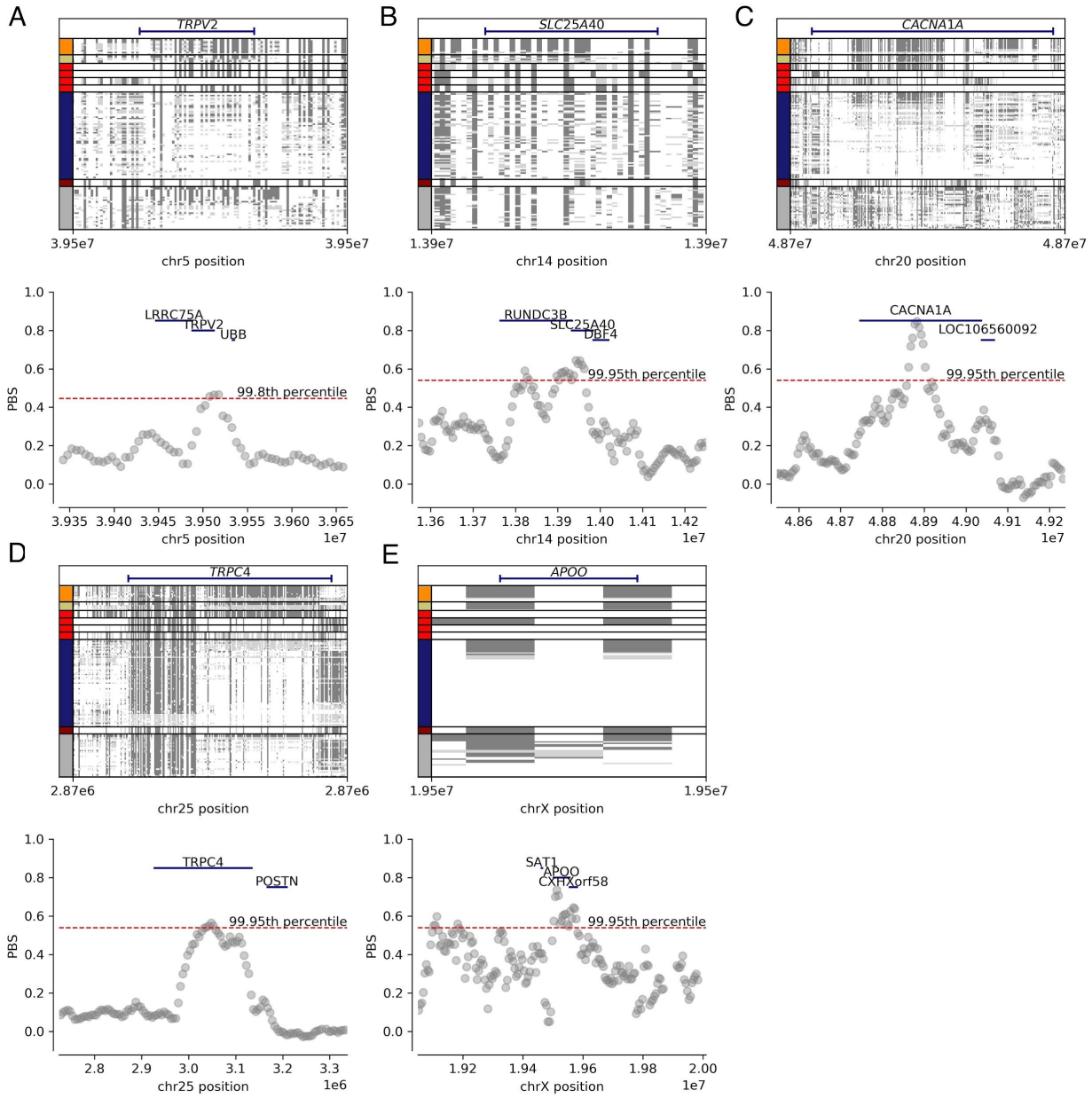


Fig. S19. Haplotype plots and PBS at outlier regions. Haplotype plots: each row represents an individual, and each column a polymorphic position in the dog genome. Cells are colored by the individual's genotype, dark gray indicates homozygous for the alternative allele, light gray indicates heterozygous positions and white indicates homozygous for the reference allele. PBS scatter plots: values in windows of 25 kilo-base pairs using a 5 kilo-base pairs slide. The percentile of the empirical distribution is shown as a red dashed horizontal line. Names of genes associated with the highest peaks are shown.

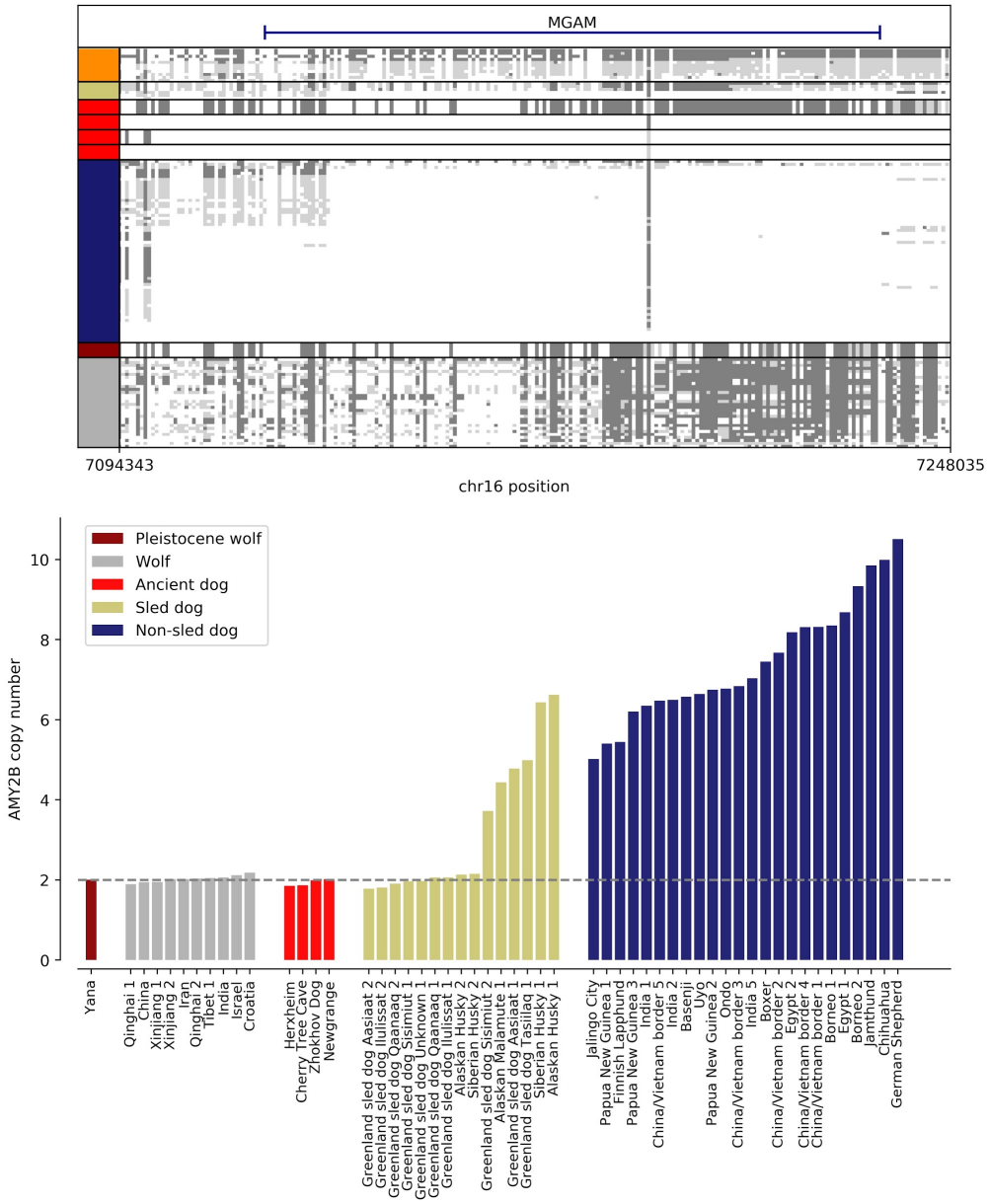


Fig. S20. MGAM and AMY2B. Above is the haplotype structure in MGAM. Each row represents an individual, and each column a polymorphic position in the dog genome. Cells are colored by the individual's genotype, dark gray indicates homozygous for the alternative allele, light gray indicates heterozygous positions and white indicates homozygous for the reference allele. The row width for ancient individuals has been increased (Zhokhov, Herxheim, Cherry Tree Cave, Newgrange and Yana). Below is the read depth based estimate of the AMY2B copy number.

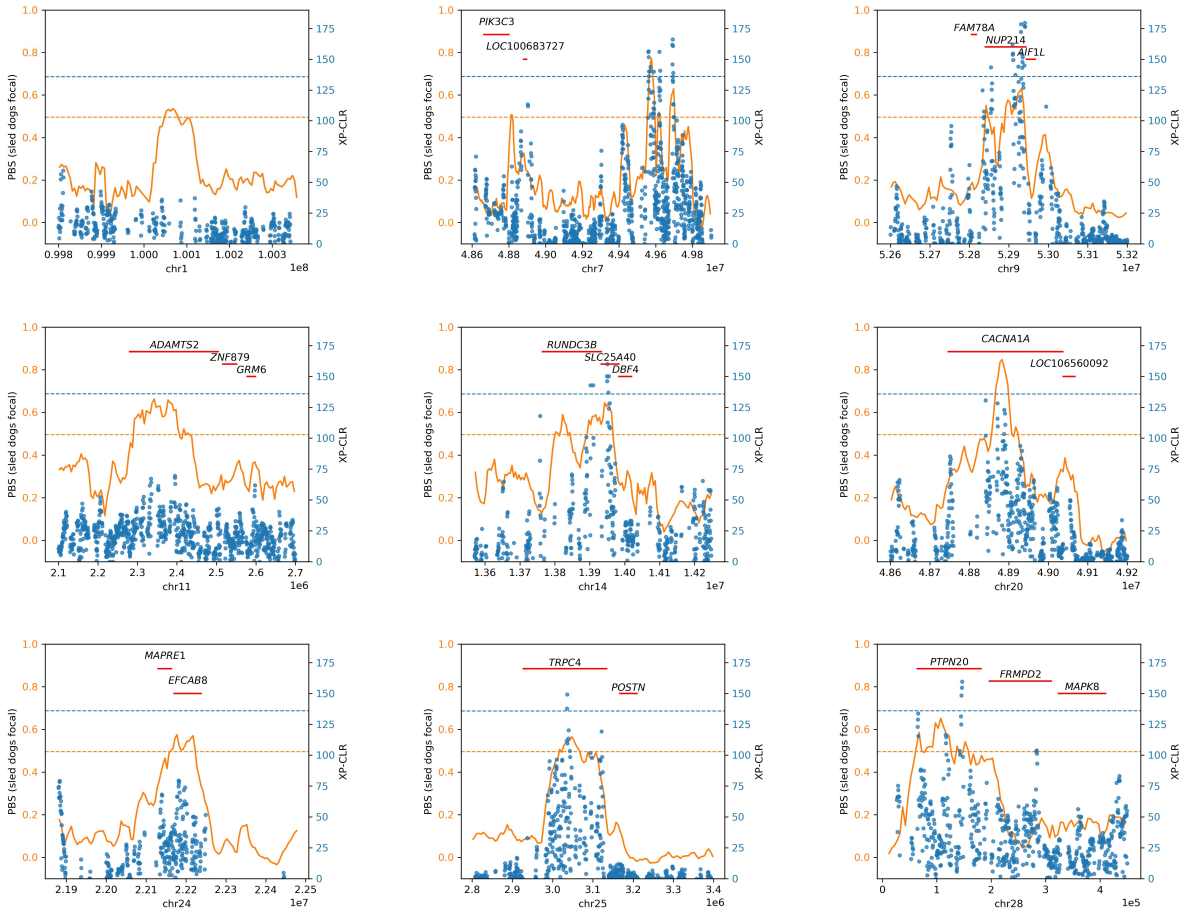


Fig. S21. XP-CLR for outlier regions in the autosomes. XP-CLR (blue) around outlier autosomal regions detected in the PBS scan (orange). Dashed lines represent the 99.9th percentile of both measures.

Supplementary Tables

Table S1. Whole genome sequencing data. Information on the whole genome sequenced samples used in this study, including the previously published samples.

Name	Publication	Type	ID	Coverage		Selected
				Wolf	Dog	
Zhokhov Dog	This study	Ancient dog	Zh-03-97 or CGG6	9.7	9.1	Treemix
Greenland sled dog Qaanaaq 1	This study	Greenland sled dog	QSON (01_05_2014)	13.6	14.0	Treemix, PBS
Greenland sled dog Qaanaaq 2	This study	Greenland sled dog	Q11 (2005)	8.5	8.9	Treemix, PBS
Greenland sled dog Aasiaat 1	This study	Greenland sled dog	Mums (71)	6.3	6.8	Treemix, PBS
Greenland sled dog Aasiaat 2	This study	Greenland sled dog	Pondus (226)	9.4	10.1	Treemix, PBS
Greenland sled dog Ilulissat 1	This study	Greenland sled dog	GS31 (1997)	6.3	4.3	Treemix, PBS
Greenland sled dog Ilulissat 2	This study	Greenland sled dog	GS16 (1997)	8.8	9.2	Treemix, PBS
Greenland sled dog Sisimiut 1	This study	Greenland sled dog	19/05/16	29.7	30.8	Treemix, PBS
Greenland sled dog Sisimiut 2	This study	Greenland sled dog	02/06/16	15.8	16.4	Treemix, PBS
Greenland sled dog Tasiilaq 1	This study	Greenland sled dog	51602 (20_08_2014)	9.1	9.4	Treemix, PBS
Greenland sled dog Tasiilaq 2	This study	Greenland sled dog	51603 (20_08_2014)	21.5	22.2	Treemix, PBS
Greenland sled dog Unknown	(*60*)	Greenland sled dog	FAMBGRD00001	13.6	13.7	PBS
Other sled dog Alaskan Malamute	(*60*)	Other sled dog	FAMBALM00001	13.6	13.8	Treemix, PBS
Other sled dog Alaskan Husky 1	(*61*)	Other sled dog	SY001	14.7	14.5	Treemix, PBS
Other sled dog Alaskan Husky 2	(*61*)	Other sled dog	SY018	12.1	11.9	Treemix, PBS
Other sled dog Siberian Husky 1	(*60*)	Other sled dog	FAMBSIH00001	11.8	11.9	Treemix, PBS
Other sled dog Siberian	(*62*)	Other sled	SiberianHusky01	58.5	57.8	Treemix, PBS

Husky 2		dog				
Other sled dog Siberian Husky 3	(*61*)	Other sled dog	KKH2684	17.4	17.1	Treemix, PBS
Greenland	(*53*)	Grey wolf	Daneborg	11.2	11.7	Treemix, PBS
Ellesmere	(*53*)	Grey wolf	GF212	7.7		Treemix, PBS
Banks Island	(*56*)	Grey wolf	SW24	15.4	15.8	Treemix, PBS
Victoria Island	(*56*)	Grey wolf	HW70	11.8		Treemix, PBS
North Baffin	(*56*)	Grey wolf	PI58	14.1	14.5	Treemix, PBS
South Baffin	(*56*)	Grey wolf	PG65	12.7	15.4	Treemix, PBS
Alaska 1	(*56*)	Grey wolf	ARF07	6.3	7.8	Treemix, PBS
Alaska 2	(*63*)	Grey wolf	AlaskanWolf	6.2		Treemix,
Mexico 1	(*7*)	Grey wolf	mxa+mxb	22.1	23.2	Treemix, PBS
Mexico 2	(*53*)	Grey wolf	MSB:Mamm:259087	2.1	2.6	Treemix, PBS
California	(*7*)	Coyote	cac	26.6		Treemix
Mexico	(*53*)	Coyote	MSB:Mamm:83943	11.3		Treemix
Portugal	(*7*)	Grey wolf	ptw	22.8	24.0	Treemix, PBS
Spain	(*7*)	Grey wolf	spw	21.3	22.4	PBS
Croatia	(*64*)	Grey wolf	RKW3919	6.2	6.0	Treemix, PBS
Israel	(*64*)	Grey wolf	RKW13759	5.5	5.2	PBS
Iran	(*7*)	Grey wolf	irw	24.7	24.7	Treemix, PBS
India	(*7*)	Grey wolf	inw	37.8	37.7	Treemix, PBS
Chinese Indigenous Dog 1	(*65*)	Dog	DogCI1	10.5	10.6	PBS
Chinese Indigenous Dog 2	(*65*)	Dog	DogCI2	9.5	10.2	Treemix, PBS
Chinese Indigenous Dog 3	(*65*)	Dog	DogCI3	7.9	7.9	PBS
Inner Mongolia 1	(*65*)	Grey wolf	GW4	7.9	7.8	Treemix, PBS
Inner Mongolia 2	(*60*)	Grey wolf	LUPZCHN00002	4.8	8.2	Treemix, PBS
Qinghai 1	(*66*)	Grey wolf	QH11* (CAN11)	22.9	22.8	Treemix, PBS
Qinghai 2	(*66*)	Grey wolf	QH16* (CAN16)	23.2	23.2	Treemix, PBS
Tibet 2	(*66*)	Grey wolf	TI32* (CAN32)	22.2	22.2	Treemix, PBS
Inner Mongolia 3	(*66*)	Grey wolf	IM06* (CAN6)	22.3	22.3	Treemix, PBS
Inner Mongolia 4	(*66*)	Grey wolf	IM07* (CAN7)	19.4	19.4	Treemix, PBS
Tibet 1	(*66*)	Grey wolf	TI09* (CAN9A)	23.0	23.0	Treemix, PBS
Shanxi 1	(*60*)	Grey wolf	LUPZCHN00006	10.5	11.6	Treemix, PBS
Shanxi 2	(*60*)	Grey wolf	LUPZCHN00005	2.7	7.0	Treemix, PBS
Xinjiang 1	(*66*)	Grey wolf	XJ24* (CAN24)	21.0	21.0	Treemix, PBS
Xinjiang 2	(*66*)	Grey wolf	XJ30* (CAN30)	23.6	23.5	Treemix, PBS
Boxer	(*67*)	Dog	canFam3.1	26.6	27.2	Treemix, PBS
German Shepherd	(*68*)	Dog	DogsGS	23.8	23.8	Treemix, PBS
Belgium Malinois	(*65*)	Dog	DogBM	9.0	9.1	Treemix, PBS

Basenji	(*64*)	Dog	RKW13764	5.3	5.2	Treemix, PBS
Dingo	(*64*)	Dog	RKW13760	5.1	4.9	Treemix, PBS
Altai	(*65*)	Grey wolf	GW1	9.6	9.6	Treemix, PBS
Chukotka	(*65*)	Grey wolf	GW2	7.7	7.7	Treemix, PBS
Bryansk	(*65*)	Grey wolf	GW3	9.8	9.8	Treemix, PBS
Tibetan Mastiff	(*65*)	Dog	DogTM	9.1	9.2	PBS
Sloughi	(*60*)	Dog	FAMBSLO00001	16.3	16.4	PBS
Finnish Lapphund	(*60*)	Dog	FAMBFIL00001	13.0	13.2	PBS
East Siberian Laika	(*60*)	Dog	FAMBESL00001	13.8	15.0	PBS
Samoyed	(*60*)	Dog	FAMBSAM00001	12.4	12.5	PBS
Swedish Lapphund	(*60*)	Dog	FAMBSWL00001	15.0	15.8	PBS
Peruvian naked	(*60*)	Dog	FAMBPEN00001	14.8	15.6	PBS
Mexican naked	(*60*)	Dog	FAMBMEN00001	13.6	14.3	PBS
Lapponian Herder	(*60*)	Dog	FAMBLAH00001	13.5	14.2	PBS
Jämthund	(*60*)	Dog	FAMBJAM00001	31.7	32.1	PBS
Gray Norwegian Elkhound	(*60*)	Dog	FAMBGNE00001	15.1	15.2	Treemix, PBS
Galgo Español	(*60*)	Dog	FAMBGAL00001	12.0	12.7	Treemix, PBS
Chihuahua	(*60*)	Dog	FAMBCHI00001	14.2	14.3	PBS
Afghan	(*60*)	Dog	FAMBAFG00001	14.4	15.3	PBS
Jalingo City	(*60*)	Dog	FAMINGR00004	14.7	14.8	PBS
Uyo	(*60*)	Dog	FAMINGR00003	11.4	11.5	Treemix, PBS
Ondo	(*60*)	Dog	FAMINGR00002	14.2	14.3	Treemix, PBS
Ibadan	(*60*)	Dog	FAMINGR00001	14.4	14.5	PBS
China/Vietnam border 1	(*60*)	Dog	FAMIVNM00001	14.5	14.6	Treemix, PBS
China/Vietnam border 2	(*60*)	Dog	FAMIVNM00002	15.1	15.2	PBS
Anhui	(*60*)	Dog	FAMICHN00025	5.4	8.0	Treemix, PBS
Yunnan 1	(*60*)	Dog	FAMICHN00023	13.4	14.3	PBS
Yunnan 2	(*60*)	Dog	FAMICHN00021	13.8	14.5	Treemix, PBS
Guizhou	(*60*)	Dog	FAMICHN00011	13.4	14.3	PBS
Guangdong	(*60*)	Dog	FAMICHN00010	15.3	16.3	PBS
Xinjiang	(*60*)	Dog	FAMICHN00019	14.7	15.6	PBS
Shaanxi 1	(*60*)	Dog	FAMICHN00017	10.2	10.6	PBS
Shaanxi 2	(*60*)	Dog	FAMICHN00016	15.0	16.3	PBS
Shanxi 1	(*60*)	Dog	FAMICHN00015	13.7	14.4	PBS
Shanxi 2	(*60*)	Dog	FAMICHN00014	14.6	14.3	PBS
Hebei	(*60*)	Dog	FAMICHN00012	15.3	16.7	PBS
Gansu 1	(*60*)	Dog	FAMICHN00006	17.3	17.1	PBS
Gansu 2	(*60*)	Dog	FAMICHN00005	13.4	14.2	PBS
Dalian	(*60*)	Dog	FAMICHN00004	13.8	15.2	PBS

Egypt 1	(*69*)	Dog	EG44	7.3	7.1	PBS
Egypt 2	(*69*)	Dog	EG49	6.6	6.5	PBS
India 1	(*69*)	Dog	ID125	11.9	11.6	PBS
India 2	(*69*)	Dog	ID137	6.7	6.6	PBS
India 3	(*69*)	Dog	ID165	9.0	8.8	PBS
India 4	(*69*)	Dog	ID168	7.7	7.6	PBS
India 5	(*69*)	Dog	ID60	12.2	11.8	PBS
India 6	(*69*)	Dog	ID91	7.4	7.2	PBS
Borneo 1	(*69*)	Dog	IN18	4.6	4.5	Treemix, PBS
Borneo 2	(*69*)	Dog	IN23	5.9	5.8	PBS
Borneo 3	(*69*)	Dog	IN29	4.9	4.8	PBS
Lebanon 1	(*69*)	Dog	LB74	6.6	6.5	PBS
Lebanon 2	(*69*)	Dog	LB79	7.8	7.7	PBS
Lebanon 3	(*69*)	Dog	LB85	7.9	7.7	PBS
Papua New Guinea 1	(*69*)	Dog	PG115	6.1	6.0	PBS
Papua New Guinea 2	(*69*)	Dog	PG122	5.7	5.6	PBS
Papua New Guinea 3	(*69*)	Dog	PG84	8.5	8.3	PBS
Qatar 1	(*69*)	Dog	QA5	6.0	6.0	PBS
Qatar 2	(*69*)	Dog	QA27	11.3	11.1	PBS
Andean Fox	(*69*)	Andean fox	Lcu2	10.5		
Taimyr	(*5*)	Ancient grey wolf	Taimyr 1	0.7		Treemix
Yana	This study	Ancient grey wolf	Y06-NP-18994 or CGG23	4.7		Treemix
Newgrange	(*40*)	Ancient dog	Newgrange	22.9		
Herxheim	(*41*)	Ancient dog	HXH	7.4		
Cherry Tree Cave	(*41*)	Ancient dog	CTC	7.7		
Port au Choix	(*3*)	Ancient dog	MU_NP50A_1	2.0		
Uyak	(*3*)	Ancient dog	HMCZ_38342	0.008		
Weyanoke Old Town 1	(*3*)	Ancient dog	DBU2-LM1	0.5		
Weyanoke Old Town 2	(*3*)	Ancient dog	DB49-LC1	0.03		
79H	(*49*)	Host	SAMEA2358416	23.7		Only CTVT work
79T	(*49*)	Tumor	SAMEA2358415	95.1		Only CTVT work
24H	(*49*)	Host	SAMEA2358414	24.5		Only CTVT work
24T	(*49*)	Tumor	SAMEA2358413	124.2		Only CTVT work

Table S2. D-statistics in figure 1D. All the D-statistics of the form D(H1, Boxer; H3, Andean Fox), computed with all dogs other than Boxer in H1, and with H3 as either the Taimyr wolf or the Yana wolf, are shown below.

H1	H3	Dstat	SE	Z score	Number of sites
Basenji	Taimyr wolf	-0.0232	0.0099	-2.33	11607
India 2	Taimyr wolf	-0.0196	0.0100	-1.96	11763
Uyo	Taimyr wolf	-0.0189	0.0095	-2.00	12841
Ibadan	Taimyr wolf	-0.0158	0.0094	-1.69	12758
India 4	Taimyr wolf	-0.0143	0.0094	-1.52	12519
Ondo	Taimyr wolf	-0.0127	0.0091	-1.39	13289
Egypt 2	Taimyr wolf	-0.0121	0.0100	-1.21	11663
India 1	Taimyr wolf	-0.0096	0.0094	-1.02	12831
Tibetan Mastiff	Taimyr wolf	-0.0080	0.0094	-0.86	13192
Samoyed	Taimyr wolf	-0.0071	0.0102	-0.70	11733
Lapponian Herder	Taimyr wolf	-0.0066	0.0099	-0.66	11701
Afghan	Taimyr wolf	-0.0057	0.0095	-0.60	12711
Jämthund	Taimyr wolf	-0.0055	0.0097	-0.57	11920
Siberian Husky 3	Taimyr wolf	-0.0054	0.0091	-0.59	13593
Peruvian naked	Taimyr wolf	-0.0050	0.0099	-0.50	11483
Papua New Guinea 3	Taimyr wolf	-0.0047	0.0104	-0.45	10302
Lebanon 3	Taimyr wolf	-0.0045	0.0105	-0.43	10266
Lebanon 1	Taimyr wolf	-0.0040	0.0109	-0.37	9243
Jalingo City	Taimyr wolf	-0.0040	0.0096	-0.41	12867
Sloughi	Taimyr wolf	-0.0038	0.0100	-0.39	11958
Xinjiang	Taimyr wolf	-0.0031	0.0093	-0.33	12788
Chinese Indigenous Dog 3	Taimyr wolf	-0.0029	0.0093	-0.31	12841
Galgo Español	Taimyr wolf	-0.0028	0.0103	-0.27	10491
Shaanxi 2	Taimyr wolf	-0.0023	0.0097	-0.24	12495
East Siberian Laika	Taimyr wolf	-0.0022	0.0098	-0.23	12532

Gansu 2	Taimyr wolf	-0.0016	0.0091	-0.18	13129
Mexican naked	Taimyr wolf	-0.0014	0.0097	-0.14	11680
Egypt 1	Taimyr wolf	-0.0009	0.0098	-0.09	11754
Guangdong	Taimyr wolf	-0.0006	0.0090	-0.07	13692
Alaskan Malamute 1	Taimyr wolf	-0.0005	0.0094	-0.06	13171
Lebanon 2	Taimyr wolf	-0.0004	0.0103	-0.04	10008
Borneo 1	Taimyr wolf	0.0004	0.0104	0.04	10640
Swedish Lapphund	Taimyr wolf	0.0005	0.0096	0.05	11550
Alaskan Husky 2	Taimyr wolf	0.0008	0.0091	0.09	13477
Shanxi 1	Taimyr wolf	0.0017	0.0099	0.17	11321
Papua New Guinea 1	Taimyr wolf	0.0018	0.0105	0.17	10260
Gansu 1	Taimyr wolf	0.0020	0.0093	0.21	13579
Dingo	Taimyr wolf	0.0021	0.0095	0.23	12563
Chihuahua	Taimyr wolf	0.0023	0.0098	0.23	11737
Chinese Indigenous Dog 1	Taimyr wolf	0.0024	0.0097	0.24	12205
India 3	Taimyr wolf	0.0025	0.0095	0.27	12323
India 6	Taimyr wolf	0.0026	0.0095	0.27	12450
Siberian Husky 2	Taimyr wolf	0.0028	0.0092	0.31	13468
Qatar 2	Taimyr wolf	0.0032	0.0096	0.33	11936
Shanxi 2	Taimyr wolf	0.0033	0.0094	0.35	12973
Gray Norwegian Elkhound	Taimyr wolf	0.0039	0.0098	0.40	11874
Borneo 3	Taimyr wolf	0.0044	0.0100	0.44	10816
China/Vietnam border 2	Taimyr wolf	0.0054	0.0091	0.60	14167
Papua New Guinea 2	Taimyr wolf	0.0055	0.0103	0.53	10779
Belgium Malinois	Taimyr wolf	0.0056	0.0100	0.56	11181
Shaanxi 1	Taimyr wolf	0.0058	0.0093	0.62	12995
Hebei	Taimyr wolf	0.0065	0.0092	0.71	13397
Greenland sled dog Aasiaat 1	Taimyr wolf	0.0065	0.0101	0.65	10716
Greenland Dog 1	Taimyr wolf	0.0066	0.0095	0.70	13111

Finnish Lapphund	Taimyr wolf	0.0066	0.0099	0.67	11734
Alaskan Husky 1	Taimyr wolf	0.0072	0.0092	0.78	13380
Qatar 1	Taimyr wolf	0.0073	0.0105	0.70	10217
Dalian	Taimyr wolf	0.0075	0.0094	0.80	12945
Yunnan 1	Taimyr wolf	0.0077	0.0091	0.84	14102
German Shepherd	Taimyr wolf	0.0089	0.0098	0.91	11712
Greenland sled dog Aasiaat 2	Taimyr wolf	0.0091	0.0094	0.96	12877
China/Vietnam border 1	Taimyr wolf	0.0097	0.0093	1.04	14143
Chinese Indigenous Dog 2	Taimyr wolf	0.0106	0.0092	1.15	13537
Guizhou	Taimyr wolf	0.0108	0.0090	1.20	13934
Greenland sled dog Ilulissat 2	Taimyr wolf	0.0116	0.0095	1.23	12455
Yunnan 2	Taimyr wolf	0.0122	0.0092	1.32	13853
Greenland sled dog Qaanaaq 2	Taimyr wolf	0.0125	0.0094	1.33	13187
Greenland sled dog Tasiilaq 1	Taimyr wolf	0.0127	0.0095	1.34	13423
Greenland sled dog Ilulissat 1	Taimyr wolf	0.0128	0.0094	1.37	12496
Greenland sled dog Qaanaaq 1	Taimyr wolf	0.0160	0.0093	1.72	13695
Anhui	Taimyr wolf	0.0162	0.0102	1.58	11380
Siberian Husky 1	Taimyr wolf	0.0169	0.0095	1.78	12398
Greenland sled dog Tasiilaq 2	Taimyr wolf	0.0172	0.0093	1.85	14023
Greenland sled dog Sisimiut 2	Taimyr wolf	0.0174	0.0092	1.89	13617
Zhokhov Dog	Taimyr wolf	0.0200	0.0090	2.22	14134
Borneo 2	Taimyr wolf	0.0211	0.0104	2.03	10649
Greenland sled dog Sisimiut 1	Taimyr wolf	0.0212	0.0092	2.32	13656
Port au Choix	Taimyr wolf	0.0699	0.0177	3.95	3374
Basenji	Yana wolf	-0.0121	0.0047	-2.54	96081
India 2	Yana wolf	-0.0021	0.0046	-0.47	97531
Uyo	Yana wolf	-0.0048	0.0046	-1.04	107147
Ibadan	Yana wolf	-0.0040	0.0045	-0.89	106990
India 4	Yana wolf	0.0061	0.0043	1.41	103897

Ondo	Yana wolf	-0.0063	0.0045	-1.40	110233
Egypt 2	Yana wolf	-0.0006	0.0045	-0.13	97263
India 1	Yana wolf	-0.0052	0.0043	-1.20	106857
Tibetan Mastiff	Yana wolf	0.0087	0.0044	1.99	109148
Samoyed	Yana wolf	0.0047	0.0046	1.01	98392
Lapponian Herder	Yana wolf	0.0062	0.0045	1.36	96844
Afghan	Yana wolf	0.0001	0.0044	0.03	107112
Jämthund	Yana wolf	0.0118	0.0046	2.58	100246
Siberian Husky 3	Yana wolf	0.0031	0.0044	0.71	113412
Peruvian naked	Yana wolf	0.0099	0.0045	2.23	96632
Papua New Guinea 3	Yana wolf	0.0028	0.0048	0.59	82501
Lebanon 3	Yana wolf	0.0153	0.0046	3.31	85311
Lebanon 1	Yana wolf	0.0132	0.0050	2.65	73772
Jalingo City	Yana wolf	0.0064	0.0044	1.45	106609
Sloughi	Yana wolf	0.0074	0.0047	1.57	98308
Xinjiang	Yana wolf	0.0187	0.0043	4.33	108121
Chinese Indigenous Dog 3	Yana wolf	0.0130	0.0043	3.00	108245
Galgo Español	Yana wolf	0.0018	0.0049	0.37	85760
Shaanxi 2	Yana wolf	0.0157	0.0043	3.65	105325
East Siberian Laika	Yana wolf	0.0149	0.0044	3.37	104267
Gansu 2	Yana wolf	0.0053	0.0044	1.21	109709
Mexican naked	Yana wolf	0.0040	0.0044	0.90	98190
Egypt 1	Yana wolf	0.0025	0.0046	0.54	96043
Guangdong	Yana wolf	0.0111	0.0043	2.59	114821
Alaskan Malamute 1	Yana wolf	0.0157	0.0044	3.55	110204
Lebanon 2	Yana wolf	0.0091	0.0047	1.91	80897
Borneo 1	Yana wolf	0.0148	0.0046	3.18	87181
Swedish Lapphund	Yana wolf	0.0159	0.0047	3.40	94748
Alaskan Husky 2	Yana wolf	0.0154	0.0044	3.48	113050

Shanxi 1	Yana wolf	0.0165	0.0044	3.71	97325
Papua New Guinea 1	Yana wolf	0.0047	0.0048	0.99	82259
Gansu 1	Yana wolf	0.0108	0.0043	2.51	113354
Dingo	Yana wolf	0.0122	0.0047	2.60	103858
Chihuahua	Yana wolf	0.0041	0.0046	0.90	99026
Chinese Indigenous Dog 1	Yana wolf	0.0090	0.0044	2.07	103201
India 3	Yana wolf	0.0086	0.0043	2.00	103004
India 6	Yana wolf	0.0046	0.0045	1.04	103040
Siberian Husky 2	Yana wolf	0.0103	0.0045	2.30	112462
Qatar 2	Yana wolf	0.0003	0.0047	0.07	97283
Shanxi 2	Yana wolf	0.0093	0.0044	2.13	108316
Gray Norwegian Elkhound	Yana wolf	0.0086	0.0046	1.88	98691
Borneo 3	Yana wolf	0.0156	0.0046	3.38	88713
China/Vietnam border 2	Yana wolf	0.0121	0.0043	2.82	117761
Papua New Guinea 2	Yana wolf	0.0123	0.0046	2.69	89325
Belgium Malinois	Yana wolf	0.0132	0.0046	2.85	92689
Shaanxi 1	Yana wolf	0.0098	0.0043	2.31	108266
Hebei	Yana wolf	0.0080	0.0042	1.91	111675
Greenland sled dog Aasiaat 1	Yana wolf	0.0199	0.0047	4.21	94696
Greenland Dog 1	Yana wolf	0.0176	0.0047	3.76	108821
Finnish Lapphund	Yana wolf	0.0111	0.0046	2.43	96292
Alaskan Husky 1	Yana wolf	0.0156	0.0044	3.51	111311
Qatar 1	Yana wolf	0.0053	0.0048	1.10	82215
Dalian	Yana wolf	0.0094	0.0043	2.19	109393
Yunnan 1	Yana wolf	0.0133	0.0042	3.16	119329
German Shepherd	Yana wolf	0.0053	0.0045	1.18	98949
Greenland sled dog Aasiaat 2	Yana wolf	0.0137	0.0045	3.02	108653
China/Vietnam border 1	Yana wolf	0.0156	0.0043	3.58	118479
Chinese Indigenous Dog 2	Yana wolf	0.0154	0.0043	3.55	112220

Guizhou	Yana wolf	0.0142	0.0042	3.36	116897
Greenland sled dog Ilulissat 2	Yana wolf	0.0177	0.0047	3.81	107350
Yunnan 2	Yana wolf	0.0126	0.0043	2.95	115891
Greenland sled dog Qaanaaq 2	Yana wolf	0.0176	0.0045	3.91	110435
Greenland sled dog Tasiilaq 1	Yana wolf	0.0182	0.0045	4.09	112636
Greenland sled dog Ilulissat 1	Yana wolf	0.0169	0.0048	3.55	103099
Greenland sled dog Qaanaaq 1	Yana wolf	0.0192	0.0045	4.25	115023
Anhui	Yana wolf	0.0254	0.0046	5.55	92926
Siberian Husky 1	Yana wolf	0.0123	0.0047	2.60	101774
Greenland sled dog Tasiilaq 2	Yana wolf	0.0171	0.0045	3.79	115980
Greenland sled dog Sisimiut 2	Yana wolf	0.0221	0.0044	4.97	113796
Zhokhov Dog	Yana wolf	0.0293	0.0044	6.62	118618
Borneo 2	Yana wolf	0.0146	0.0049	2.95	83668
Greenland sled dog Sisimiut 1	Yana wolf	0.0180	0.0046	3.89	114290
Port au Choix	Yana wolf	0.0636	0.0069	9.20	27453

Table S3. Data figure S8

ID	Type	Publication	Data type
5128	Village Dog India-West Bengal	(*70*)	snp data
8277	Village Dog US-Alaska	(*70*)	snp data
8278	Village Dog US-Alaska	(*70*)	snp data
8279	Village Dog US-Alaska	(*70*)	snp data
AlaskanHusky_SY001	Alaskan Husky 1	(*61*)	genome
AlaskanHusky_SY018	Alaskan Husky 1	(*61*)	genome
AMAL_14509	Alaskan Malamute	(*62*)	snp data
AMAL_27127	Alaskan Malamute	(*62*)	snp data
AMAL_29540	Alaskan Malamute	(*62*)	snp data
AMAL_29541	Alaskan Malamute	(*62*)	snp data
AMAL_29542	Alaskan Malamute	(*62*)	snp data
AMAL_29543	Alaskan Malamute	(*62*)	snp data
AMAL_29544	Alaskan Malamute	(*62*)	snp data
AMAL_29545	Alaskan Malamute	(*62*)	snp data
AMAL_29546	Alaskan Malamute	(*62*)	snp data
AMAL_29547	Alaskan Malamute	(*62*)	snp data
canFam3.1	Boxer	(*67*)	genome
Box_LU128	Boxer	(*70*)	snp data
Box_LU132	Boxer	(*70*)	snp data
BSJI_12301	Basenji	(*70*)	snp data
BSJI_35551	Basenji	(*70*)	snp data
BSJI_35554	Basenji	(*70*)	snp data
BSJI_45738	Basenji	(*70*)	snp data
BSJI_51967	Basenji	(*70*)	snp data
CGG6	Ancient dog	This study	genome
EG41	Village Dog Egypt	(*70*)	snp data
EG43	Village Dog Egypt	(*70*)	snp data
EG44	Village Dog Egypt	(*69*)	genome
EG49	Village Dog Egypt	(*69*)	genome
Mums (71)	Greenland sled dog	This study	genome

Pondus (226)	Greenland sled dog	This study	genome
GShepDog	German Shepherd Dog	(*65*)	genome
GSh_GT200	German Shepherd Dog	(*70*)	snp data
GSh_GT205	German Shepherd Dog	(*70*)	snp data
GSh_GT208	German Shepherd Dog	(*70*)	snp data
GSl_GT226	Greenland sled dog	(*70*)	snp data
GSl_GT227	Greenland sled dog	(*70*)	snp data
GSl_GT228	Greenland sled dog	(*70*)	snp data
GSl_GT229	Greenland sled dog	(*70*)	snp data
GSl_GT230	Greenland sled dog	(*70*)	snp data
GSl_GT231	Greenland sled dog	(*70*)	snp data
GSl_GT232	Greenland sled dog	(*70*)	snp data
GSl_GT233	Greenland sled dog	(*70*)	snp data
GSl_GT234	Greenland sled dog	(*70*)	snp data
GSl_GT235	Greenland sled dog	(*70*)	snp data
GSl_GT236	Greenland sled dog	(*70*)	snp data
GSl_GT237	Greenland sled dog	(*70*)	snp data
H5-HUC12-MALE-N/A	Boxer	(*70*)	snp data
HUSK_11480	Siberian Husky	(*62*)	snp data
HUSK_11487	Siberian Husky	(*62*)	snp data
HUSK_11490	Siberian Husky	(*62*)	snp data
HUSK_14529	Siberian Husky	(*62*)	snp data
HUSK_15872	Siberian Husky	(*62*)	snp data
HUSK_15873	Siberian Husky	(*62*)	snp data
HUSK_28487	Siberian Husky	(*62*)	snp data
HUSK_28925	Siberian Husky	(*62*)	snp data
HUSK_28926	Siberian Husky	(*62*)	snp data
HUSK_29464	Siberian Husky	(*62*)	snp data
ID15	Village Dog India	(*70*)	snp data
ID21	Village Dog India	(*70*)	snp data
ID58	Village Dog India	(*70*)	snp data
ID60	Village Dog India	(*69*)	genome

ID104	Village Dog India	(*70*)	snp data
ID149	Village Dog India	(*70*)	snp data
IN1	Village Dog Borneo	(*70*)	snp data
IN4	Village Dog Borneo	(*70*)	snp data
IN8	Village Dog Borneo	(*70*)	snp data
IN18	Village Dog Borneo	(*69*)	genome
IN20	Village Dog Borneo	(*70*)	snp data
IN23	Village Dog Borneo	(*69*)	genome
IN30	Village Dog Borneo	(*70*)	snp data
IN35	Village Dog Borneo	(*70*)	snp data
IN44	Village Dog Java	(*70*)	snp data
IN45	Village Dog Java	(*70*)	snp data
PFZ1A08	Village Dog US-Alaska	(*70*)	snp data
PFZ1A09	Village Dog US-Alaska	(*70*)	snp data
PFZ1B01	Village Dog US-Alaska	(*70*)	snp data
PFZ1B02	Village Dog Peru	(*70*)	snp data
PFZ1B05	Village Dog India	(*70*)	snp data
PFZ1B11	Village Dog Peru	(*70*)	snp data
PFZ1B12	Village Dog Peru	(*70*)	snp data
PFZ1C05	Village Dog Vietnam	(*70*)	snp data
PFZ1D08	Village Dog US-Alaska	(*70*)	snp data
PFZ1E05	Afghan Hound	(*70*)	snp data
PFZ1F04	Village Dog Papua New Guinea	(*70*)	snp data
PFZ1F11	Village Dog Peru	(*70*)	snp data
PFZ1G06	Village Dog US-Alaska	(*70*)	snp data
PFZ1H07	Village Dog US-Alaska	(*70*)	snp data
PFZ2A03	Village Dog Vietnam	(*70*)	snp data
PFZ2B08	Village Dog US-Alaska	(*70*)	snp data
PFZ2D02	Village Dog Indonesia	(*70*)	snp data
PFZ2D04	Basenji	(*70*)	snp data
PFZ2E05	Village Dog US-Alaska	(*70*)	snp data
PFZ2F06	Village Dog Peru	(*70*)	snp data

PFZ2G10	Village Dog Peru	(*70*)	snp data
PFZ2H12	Village Dog Peru	(*70*)	snp data
PFZ3A06	Village Dog Mexico	(*70*)	snp data
PFZ3C05	Village Dog Mexico	(*70*)	snp data
PFZ3C06	Village Dog Mexico	(*70*)	snp data
PFZ3D11	Village Dog India	(*70*)	snp data
PFZ3E06	Village Dog Mexico	(*70*)	snp data
PFZ3F06	Village Dog Mexico	(*70*)	snp data
PFZ3G06	Village Dog Mexico	(*70*)	snp data
PFZ3H05	Village Dog Mexico	(*70*)	snp data
PFZ4A08	German Shepherd Dog	(*70*)	snp data
PFZ4B05	Village Dog Borneo	(*70*)	snp data
PFZ4D01	Chinese Crested	(*70*)	snp data
PFZ5A11	Afghan Hound	(*70*)	snp data
PFZ5B11	Afghan Hound	(*70*)	snp data
PFZ5C08	Afghan Hound	(*70*)	snp data
PFZ5C12	Chinese Crested	(*70*)	snp data
PFZ5G08	Village Dog Papua New Guinea	(*70*)	snp data
PFZ5H07	German Shepherd Dog	(*70*)	snp data
PFZ5H08	Village Dog Vietnam	(*70*)	snp data
PFZ6A04	Alaskan Malamute	(*70*)	snp data
PFZ6B11	Alaskan Malamute	(*70*)	snp data
PFZ6C01	Chinese Crested	(*70*)	snp data
PFZ6D07	Alaskan Malamute	(*70*)	snp data
PFZ6E07	Alaskan Malamute	(*70*)	snp data
PFZ6E10	Mastiff	(*70*)	snp data
PFZ6F11	Alaskan Malamute	(*70*)	snp data
PFZ6G01	Mastiff	(*70*)	snp data
PFZ6G03	Mastiff	(*70*)	snp data
PFZ6G07	Alaskan Malamute	(*70*)	snp data
PFZ6H03	Mastiff	(*70*)	snp data
PFZ6H07	Alaskan Malamute	(*70*)	snp data

PFZ6H10	Alaskan Malamute	(*70*)	snp data
PFZ7C02	Alaskan Malamute	(*70*)	snp data
PFZ7H12	Alaskan Malamute	(*70*)	snp data
PFZ8B09	Samoyed	(*70*)	snp data
PFZ9A02	Samoyed	(*70*)	snp data
PFZ9B09	Chinese Shar-pei	(*70*)	snp data
PFZ9B12	Belgian Malinois	(*70*)	snp data
PFZ9C09	Samoyed	(*70*)	snp data
PFZ9C11	Alaskan Malamute	(*70*)	snp data
PFZ9D05	Samoyed	(*70*)	snp data
PFZ9D06	Alaskan Malamute	(*70*)	snp data
PFZ9D10	Belgian Malinois	(*70*)	snp data
PFZ9E06	Chinese Shar-pei	(*70*)	snp data
PFZ9F12	Mastiff	(*70*)	snp data
PFZ9G11	Belgian Malinois	(*70*)	snp data
PFZ9H11	Mastiff	(*70*)	snp data
PFZ10B10	Chihuahua	(*70*)	snp data
PFZ10C09	Siberian Husky	(*70*)	snp data
PFZ10D03	Chihuahua	(*70*)	snp data
PFZ10F03	Chihuahua	(*70*)	snp data
PFZ12D02	German Shepherd Dog	(*70*)	snp data
PFZ14G08	Chihuahua	(*70*)	snp data
PFZ15A05	Chihuahua	(*70*)	snp data
PFZ15C11	Siberian Husky	(*70*)	snp data
PFZ15H03	Basenji	(*70*)	snp data
PFZ16A02	Afghan Hound	(*70*)	snp data
PFZ16A10	Samoyed	(*70*)	snp data
PFZ16B12	Samoyed	(*70*)	snp data
PFZ16F06	Siberian Husky	(*70*)	snp data
PFZ16G03	Siberian Husky	(*70*)	snp data
PFZ17A12	Siberian Husky	(*70*)	snp data
PFZ17B08	Samoyed	(*70*)	snp data

PFZ17B12	Afghan Hound	(*70*)	snp data
PFZ17C09	Siberian Husky	(*70*)	snp data
PFZ17D05	Siberian Husky	(*70*)	snp data
PFZ17E08	Siberian Husky	(*70*)	snp data
PFZ17E12	Siberian Husky	(*70*)	snp data
PFZ17F09	Afghan Hound	(*70*)	snp data
PFZ17G01	Samoyed	(*70*)	snp data
PFZ17G09	Siberian Husky	(*70*)	snp data
PFZ18A03	Belgian Malinois	(*70*)	snp data
PFZ22G08	Siberian Husky	(*70*)	snp data
PFZ23H11	Boxer	(*70*)	snp data
PFZ30B03	Village Dog Peru	(*70*)	snp data
PFZ31B05	Village Dog Peru	(*70*)	snp data
PFZ33F06	Boxer	(*70*)	snp data
PFZ34B10	Boxer	(*70*)	snp data
PFZ35A12	Mastiff	(*70*)	snp data
PFZ35D12	Mastiff	(*70*)	snp data
PFZ35G01	Chihuahua	(*70*)	snp data
PFZ35G11	Mastiff	(*70*)	snp data
PFZ35H05	German Shepherd Dog	(*70*)	snp data
PFZ36B07	Siberian Husky	(*70*)	snp data
PFZ37B12	Boxer	(*70*)	snp data
PFZ38H04	Boxer	(*70*)	snp data
PFZ39E06	Belgian Malinois	(*70*)	snp data
PFZ40F08	German Shepherd Dog	(*70*)	snp data
PFZ40G07	German Shepherd Dog	(*70*)	snp data
PFZ41C04	German Shepherd Dog	(*70*)	snp data
PFZ41F04	Boxer	(*70*)	snp data
PFZ43B01	Neapolitan Mastiff	(*70*)	snp data
PFZ43C02	Siberian Husky	(*70*)	snp data
PFZ43E11	Chihuahua	(*70*)	snp data
PFZ44B11	Siberian Husky	(*70*)	snp data

PFZ44C03	Siberian Husky	(*70*)	snp data
PFZ44F10	Siberian Husky	(*70*)	snp data
PFZ44F12	Siberian Husky	(*70*)	snp data
PG40	Village Dog Papua New Guinea	(*70*)	snp data
PG49	Village Dog Papua New Guinea	(*70*)	snp data
PG83	Village Dog Papua New Guinea	(*70*)	snp data
PG115	Village Dog Papua New Guinea	(*69*)	genome
PG116	Village Dog Papua New Guinea	(*70*)	snp data
QSON (01_05_2014)	Greenland sled dog	This study	genome
FAMBALM00001	Alaskan Malamute	(*60*)	genome
FAMBSIH00001	Siberian Husky	(*60*)	genome
ShP_GT358	Chinese Shar-pei	(*70*)	snp data
ShP_GT359	Chinese Shar-pei	(*70*)	snp data
ShP_GT366	Chinese Shar-pei	(*70*)	snp data
KKH2684	Siberian Husky	(*61*)	genome
51602 (20_08_2014)	Greenland sled dog	This study	genome
VN27	Village Dog Vietnam	(*70*)	snp data
VN32	Village Dog Vietnam	(*70*)	snp data
VN50	Village Dog Vietnam	(*70*)	snp data
VN73	Village Dog Vietnam	(*70*)	snp data

Table S4

Regions above the 99.5th and 99.95th percentiles of the PBS empirical distribution. Overlapping windows were merged into single regions. Regions below the 0.05th percentile are also shown (see Table S5 for a specific test scanning regions highly differentiated in other dogs relative to sled dogs and wolves).

chr	start	end	pbs	percentile	overlapping genes
chr1	14580000	14680000	0.335	99.5th	KIAA1468
chr1	37360000	37460000	0.329	99.5th	GRM1
chr1	47800000	47900000	0.340	99.5th	TULP4
chr1	51920000	52040000	0.343	99.5th	QKI
chr1	53700000	53840000	0.342	99.5th	PDE10A
chr1	54440000	54560000	0.352	99.5th	RPS6KA2
chr1	63680000	63840000	0.347	99.5th	NKAIN2
chr1	68980000	69100000	0.379	99.5th	
chr1	69240000	69360000	0.361	99.5th	EPB41L2
chr1	80000000	80100000	-0.131	0.05th	
chr1	85540000	85640000	0.329	99.5th	GDA
chr1	98080000	98260000	0.355	99.5th	PHF2
chr1	100000000	100160000	0.481	99.95th	
chr1	100520000	100740000	0.353	99.5th	ZNF304
chr1	104560000	104680000	0.427	99.5th	
chr3	19260000	19360000	0.333	99.5th	
chr3	21860000	21960000	0.348	99.5th	
chr3	36960000	37080000	0.336	99.5th	CHRNA7
chr3	40780000	40900000	0.361	99.5th	ADAMTS17
chr3	43440000	43540000	0.343	99.5th	
chr4	23960000	24180000	0.449	99.5th	ANXA7, MSS51, PPP3CB, USP54, MYOZ1, SYNPO2L
chr4	24220000	24320000	0.331	99.5th	ZSWIM8, NDST2, CAMK2G
chr4	32800000	32940000	0.374	99.5th	
chr4	64060000	64180000	0.354	99.5th	PARP8
chr4	72020000	72180000	0.413	99.5th	
chr4	74100000	74280000	0.379	99.5th	ADAMTS12, TARS

chr4	79480000	79800000	0.397	99.5th	
chr5	1680000	1800000	0.340	99.5th	
chr5	9300000	9480000	0.404	99.5th	PKNOX2, TMEM218, SLC37A2, CCDC15
chr5	33080000	33360000	0.369	99.5th	RANGRF, SLC25A35, ARHGEF15, ODF4, KRBA2, RPL26, RNF222, NDEL1, MYH10
chr5	38320000	38420000	0.356	99.5th	
chr5	38600000	38740000	0.354	99.5th	
chr5	39480000	39600000	0.330	99.5th	TRPV2, LRRC75A, UBB, CENPV
chr5	41220000	41360000	0.428	99.5th	SHMT1, TOP3A, MIEF2, FLII, LLGL1, ALKBH5
chr5	60760000	60960000	0.389	99.5th	CAMTA1
chr5	61600000	61780000	0.400	99.5th	ERRFI1
chr5	62480000	62640000	0.388	99.5th	H6PD
chr5	63760000	63880000	0.360	99.5th	SPIRE2, FANCA, ZNF276, VPS9D1, SPATA2L, CDK10
chr5	72420000	72580000	0.336	99.5th	WWOX
chr6	6540000	6680000	0.365	99.5th	ABHD11, STX1A, BUD23, DNAJC30, VPS37D, MLXIPL, TBL2
chr6	12600000	12760000	0.459	99.5th	TNRC18, SLC29A4, WIPI2
chr6	28860000	29040000	0.417	99.5th	MKL2
chr6	31660000	31760000	0.329	99.5th	CLEC16A, DEXI
chr6	46900000	47100000	-0.169	0.05th	RNPC3 (AMY2B)
chr6	55700000	56000000	0.446	99.5th	DR1, CCDC18, TMED5, MTF2
chr6	56160000	56420000	0.426	99.5th	FAM69A, RPL5, EVI5
chr6	67520000	67660000	0.377	99.5th	
chr6	68500000	68660000	0.352	99.5th	
chr6	68920000	69040000	0.352	99.5th	DNAJB4, FUBP1, NEXN, MIGA1
chr7	3180000	3320000	0.333	99.5th	
chr7	34820000	34940000	0.369	99.5th	AKT3
chr7	48800000	48900000	0.355	99.5th	PIK3C3
chr7	49480000	49820000	0.532	99.95th	
chr8	1460000	1560000	0.375	99.5th	

chr9	5800000	5900000	0.332	99.5th	CD300E
chr9	23620000	23740000	0.354	99.5th	SRCIN1, ARHGAP23
chr9	52820000	53000000	0.516	99.95th	NUP214, AIF1L, LAMC3
chr10	61860000	61980000	0.350	99.5th	CCT4, COMMD1
chr11	2060000	2180000	0.349	99.5th	RUFY1
chr11	2200000	2500000	0.621	99.95th	ADAMTS2
chr11	4180000	4300000	0.347	99.5th	KCNN2
chr11	5100000	5280000	0.421	99.5th	FEM1C
chr11	14740000	14860000	0.365	99.5th	
chr11	25940000	26100000	0.407	99.5th	CDC25C, FAM53C, KDM3B, REEP2, EGR1, ETF1
chr11	39880000	39980000	0.346	99.5th	MLLT3
chr11	62400000	62540000	0.375	99.5th	ZNF462
chr12	25840000	25980000	0.390	99.5th	KHDRBS2
chr12	69260000	69460000	0.446	99.5th	
chr13	42980000	43080000	0.329	99.5th	GABRB1
chr14	11880000	12060000	0.451	99.5th	WASL, LMOD2, ASB15, NDUFA5
chr14	12480000	12680000	0.473	99.95th	
chr14	13140000	13240000	0.347	99.5th	KIAA1324L
chr14	13780000	14040000	0.549	99.95th	SLC25A40, ABCB1, RUNDC3B, DBF4, ADAM22
chr14	18300000	18400000	0.331	99.5th	CDK6
chr14	19300000	19420000	0.359	99.5th	GNGT1, TFPI2
chr14	21620000	21800000	0.405	99.5th	
chr15	16980000	17100000	0.335	99.5th	EBNA1BP2, FAM183A
chr15	23040000	23160000	0.348	99.5th	MYF6, MYF5, LIN7A
chr15	33880000	34080000	0.422	99.5th	CRADD
chr16	4820000	4960000	-0.154	0.05th	
chr16	7120000	7440000	-0.153	0.05th	MGAM, TAS2R38, CLEC5A, COR9A7, TAS2R5, TAS2R4, TAS2R3
chr16	16740000	16840000	-0.127	0.05th	
chr16	39180000	39280000	0.361	99.5th	
chr16	39600000	39780000	0.389	99.5th	

chr17	18580000	18680000	0.331	99.5th	FAM228B, ITSN2
chr17	18960000	19140000	0.354	99.5th	NCOA1, CENPO, PTRHD1
chr17	52220000	52400000	0.395	99.5th	TRIM33, BCAS2, DENND2C, AMPD1
chr18	6360000	6580000	0.446	99.5th	VOPP1, BLVRA, COA1
chr18	7120000	7220000	0.340	99.5th	MRPL32, PSMA2
chr19	3060000	3160000	0.340	99.5th	MAML3, MGST2
chr19	52240000	52380000	0.350	99.5th	
chr20	34120000	34420000	0.473	99.95th	ERC2
chr20	43780000	43940000	0.446	99.5th	GATAD2A, MAU2, SUGP1, TM6SF2, HAPLN4
chr20	48740000	48980000	0.614	99.95th	CACNA1A
chr20	49600000	49720000	0.346	99.5th	
chr20	55400000	55560000	0.416	99.5th	ANKRD24, SIRT6, CREB3L3, MAP2K2, ZBTB7A, PIAS4
chr21	22100000	22260000	0.399	99.5th	EMSY
chr21	22280000	22400000	0.362	99.5th	EMSY, THAP12
chr21	24000000	24160000	0.374	99.5th	P4HA3, PPME1, C2CD3
chr21	27320000	27440000	0.341	99.5th	COR51A16, MMP26
chr22	1000000	1260000	0.338	99.5th	RNASEH2B
chr22	1280000	1460000	0.336	99.5th	
chr22	1680000	1940000	0.380	99.5th	KCNRG, TRIM13, SPRYD7
chr23	2980000	3140000	0.453	99.5th	
chr23	21220000	21340000	0.378	99.5th	ZNF385D
chr23	42800000	42960000	0.391	99.5th	
chr24	20000	300000	0.439	99.5th	CST8, CST11, CSTL1, NAPB, GZF1, NXT1
chr24	2120000	2260000	0.348	99.5th	XRN2, KIZ
chr24	3000000	3160000	0.383	99.5th	INSM1, CFAP61
chr24	5160000	5320000	0.422	99.5th	BANF2
chr24	6000000	6200000	0.472	99.5th	OTOR, SNRPB2
chr24	7820000	7940000	0.353	99.5th	MACROD2
chr24	13340000	13480000	0.403	99.5th	PLCB1
chr24	16360000	16500000	0.357	99.5th	PROKR2, CDS2, PCNA, TMEM230

chr24	21280000	21480000	0.436	99.5th	MYLK2, FOXS1, DUSP15, TTLL9, PDRG1, XKR7, CCM2L, HCK
chr24	22080000	22280000	0.503	99.95th	DNMT3B, MAPRE1, EFCAB8, SUN5, BPIFB2
chr25	620000	800000	0.349	99.5th	FOXO1
chr25	2960000	3160000	0.508	99.95th	TRPC4
chr25	30380000	30480000	0.348	99.5th	
chr27	4280000	4440000	0.366	99.5th	LARP4, FAM186A
chr27	4560000	4900000	0.461	99.5th	LIMA1, CERS5, COX14, GPD1, SMARCD1, ASIC1, RACGAP1, AQP5, AQP2, FAIM2, BCDIN3D, NCKAP5L
chr27	5000000	5180000	0.432	99.5th	FMNL3, PRPF40B, FAM186B, MCRS1, KCNH3, SPATS2
chr27	40960000	41140000	0.417	99.5th	PARP11, CRACR2A
chr28	20000	240000	0.535	99.95th	PTPN20, FRMPD2
chr28	800000	920000	0.341	99.5th	WDFY4, LRRC18
chr28	1880000	2080000	0.411	99.5th	WASHC2C, ZFAND4, MARCH8
chr28	2240000	2340000	0.333	99.5th	OR13A1, COR6D7
chr28	11120000	11260000	0.361	99.5th	CRTAC1
chr28	15520000	15640000	0.376	99.5th	NT5C2
chr30	400000	760000	0.386	99.5th	LPCAT4, NUTM1
chr30	18520000	18700000	0.392	99.5th	
chr30	33500000	33600000	0.334	99.5th	
chr31	1220000	1320000	0.335	99.5th	
chr31	3200000	3400000	0.418	99.5th	
chr31	3760000	3880000	0.368	99.5th	
chr33	3520000	3620000	0.334	99.5th	
chr33	13680000	13840000	0.386	99.5th	MYH15, CIP2A
chr33	26240000	26360000	0.376	99.5th	SEC22A, ADCY5
chr33	31260000	31360000	0.339	99.5th	ATP13A3, LRRC15, CPN2
chr34	12600000	12760000	0.415	99.5th	PIK3CA, KCNMB3
chr35	14460000	14640000	0.426	99.5th	JARID2
chr36	19660000	19840000	0.413	99.5th	LNPK
chrX	14520000	14620000	0.339	99.5th	CDKL5

chrX	15900000	16160000	-0.122	0.05th	RPS6KA3, BCLAF3, MAP7D2, EIF1AX
chrX	19060000	19720000	0.590	99.95th	APOO, PTCHD1, PRDX4, ACOT9, SAT1, CXHXorf58, KLHL15, EIF2S3
chrX	23960000	24140000	0.391	99.5th	IL1RAPL1
chrX	28740000	28860000	-0.122	0.05th	
chrX	38520000	38640000	0.353	99.5th	
chrX	41500000	41620000	0.359	99.5th	ZNF182
chrX	43580000	43740000	0.375	99.5th	SHROOM4

Table S5

Regions above the 99.95th percentiles of the PBS empirical distribution using other dogs as the focal population. Overlapping windows were merged into single regions. We note AMY2B is not annotated in the chromosome 6 of canFam3.1, but it is immediately upstream of RNPC3 (*19*).

chr	start	end	pbs	percentile	overlapping genes
chr6	46880000	47380000	1.326	99.95th	RNPC3 (AMY2B)
chr16	7120000	7360000	0.918	99.95th	MGAM, TAS2R38, CLEC5A, COR9A7
chrX	15940000	16120000	0.719	99.95th	RPS6KA3, MAP7D2, EIF1AX

Table S6

Interval enrichment analysis based on the regions in the top 99.95 and 99.5 percentile of the empirical distribution of the PBS, with the sled dogs as the focal population.

GO term	GO description	Nr of genes in term	Nr intervals with term	Empirical P-value	Corrected P-value
PBS threshold: 0.451 (99.95th percentile), 14 intervals					
GO:0014051	gamma-aminobutyric acid secretion	3	2	1.9996×10^{-4}	0.119
GO:0070509	calcium ion import	14	2	1.9996×10^{-4}	0.119
GO:0070588	calcium ion transmembrane transport	55	2	3.9992×10^{-3}	0.382
PBS threshold: 0.328 (99.5th percentile), 113 intervals					
GO:0000122	negative regulation of transcription by RNA polymerase II	434	13	0.0115977	0.9998
GO:0001502	cartilage condensation	8	2	0.00219956	0.95041
GO:0004693	cyclin-dependent protein serine/threonine kinase activity	19	2	0.0469906	1
GO:0006396	RNA processing	22	2	0.0467906	1
GO:0006397	mRNA processing	24	3	0.0103979	0.9998
GO:0006406	mRNA export from nucleus	22	2	0.0475905	1
GO:0006471	protein ADP-ribosylation	6	2	0.00519896	0.995001

GO:0006497	protein lipidation	5	2	0.00279944	0.973205
GO:0007274	neuromuscular synaptic transmission	19	2	0.0289942	1
GO:0009267	cellular response to starvation	35	3	0.0253949	1
GO:0010628	positive regulation of gene expression	106	5	0.0195961	1
GO:0010807	regulation of synaptic vesicle priming	3	2	0.00579884	0.996401
GO:0014051	gamma-aminobutyric acid secretion	3	2	0.00459908	0.992601
GO:0015031	protein transport	50	3	0.0465907	1
GO:0017124	SH3 domain binding	30	3	0.0327934	1
GO:0018401	peptidyl-proline hydroxylation to 4-hydroxy-L-proline	8	2	0.0139972	1
GO:0019213	deacetylase activity	6	2	0.00319936	0.982404
GO:0030855	epithelial cell differentiation	31	2	0.0495901	1
GO:0030968	endoplasmic reticulum unfolded protein response	36	3	0.0239952	1
GO:0032454	histone demethylase activity (H3-K9 specific)	10	2	0.0293941	1
GO:0032870	cellular response to hormone stimulus	19	2	0.0277944	1
GO:0033169	histone H3-K9 demethylation	8	2	0.0127974	1
GO:0034497	protein localization to phagophore assembly site	8	2	0.0039992	0.989202
GO:0035035	histone acetyltransferase binding	12	2	0.0335933	1
GO:0035914	skeletal muscle cell differentiation	26	2	0.0235953	1
GO:0042149	cellular response to glucose starvation	18	2	0.0367926	1
GO:0042169	SH2 domain binding	16	2	0.0479904	1
GO:0042273	ribosomal large subunit biogenesis	13	2	0.0209958	1

GO:0045159	myosin II binding	6	2	0.00659868	0.9982
GO:0045773	positive regulation of axon extension	11	2	0.0113977	0.9998
GO:0048015	phosphatidylinositol-mediated signaling	23	2	0.0485903	1
GO:0048146	positive regulation of fibroblast proliferation	22	2	0.0475905	1
GO:0051721	protein phosphatase 2A binding	14	2	0.020196	1
GO:0070509	calcium ion import	14	2	0.039992	1
GO:0070588	calcium ion transmembrane transport	55	4	0.0253949	1
GO:0080025	phosphatidylinositol-3,5-bisphosphate binding	12	2	0.0153969	1
GO:1901098	positive regulation of autophagosome maturation	5	2	0.00139972	0.883623

References